03/10/00 JC784 U.S. PTO

03-13-00

Box Seq A

PATENT

Attorney Docket No. SALK1510-3

___ NEW PATENT APPLICATION

X CONTINUATION-IN-PART

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL" MAILING LABEL NO EL476992837US

DATE OF DEPOSIT March 10, 2000

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 C F R 1 10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT COMMISSIONER FOR PATENTS, WASHINGTON, D C 20231

Lynn Morkunas

(TYPED OR PRINTED NAME OF PERSON MAILING PAPER OR FEE)

(SIGNATURE OF PERSON MAILING PAPER OR FEE)

JC678 U.S. PTO 09/522753 03/10/00

ASSISTANT COMMISSIONER
FOR PATENTS
Box Patent Application
Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the new patent application of

Inventors: Ronald M. Evans, J. Don Chen and Peter Ordentlich

For: **A FAMILY OF TRANSCRIPTIONAL CO-REPRESSORS THAT INTERACT WITH NUCLEAR HORMONE RECEPTORS AND USES THEREFOR**

This is a request for filing a continuation-in-part under 35 U.S.C. 111(A) and 37 C.F.R. 1.53(b), of U.S. Application Serial No. 08/522,726, filed September 1, 1995, now pending.

Enclosed are:

X 75 pages of the Specification, which includes 7 pages of the claims and 1 page of the Abstract;

X 12 sheets of drawing(s) ___ Formal; X Informal;

X A Declaration (unexecuted);

X 67-Page Sequence Listing;

X computer readable disk containing Sequence Listing; and

X Statement Under 37 C.F.R. §§1.821(f) and (g).

FULL NAME OF FIRST INVENTOR	*LAST NAME:* EVANS	*FIRST NAME:* RONALD	*MIDDLE NAME:* M.
CITIZENSHIP	*STATE OR FOREIGN COUNTRY:* US		
POST OFFICE ADDRESS	*POST OFFICE ADDRESS:* 1471 Cottontail Lane	*CITY AND STATE:* La Jolla, California	*ZIP CODE:* 92037
FULL NAME OF SECOND INVENTOR	*LAST NAME:* CHEN	*FIRST NAME:* J.	*MIDDLE NAME:* DON
CITIZENSHIP	*STATE OR FOREIGN COUNTRY:* Taiwan		
POST OFFICE ADDRESS	*POST OFFICE ADDRESS:* 7548 Charmant Drive, #1416	*CITY AND STATE:* San Diego, California	*ZIP CODE:* 92126

FULL NAME OF SECOND INVENTOR	*LAST NAME:* ORDENTLICH	*FIRST NAME:* PETER.	*MIDDLE NAME:*
CITIZENSHIP	*STATE OR FOREIGN COUNTRY:*		
POST OFFICE ADDRESS	*POST OFFICE ADDRESS:*	*CITY AND STATE:*	*ZIP CODE:*

The filing fee has been calculated as shown below:

For	Number Filed		Number Extra		Rate			Fee	
					Small Entity	Other Entity		Small Entity	Other Entity
Total Claims		=		X	$ 9	$18	=	$ 0	$.00
Independent Claims		=		X	$39	$78	=	$ 0	$.00
Multiple Dependent Claims Presented: ___ Yes _X_ No					$130	$260			$.00
				BASIC FEE	$345	$690		$ 0	$.00
					TOTAL FEE			$ 0	$.00

_X_ The payment of the filing fee is to be deferred until the Declaration is filed. Do not charge our deposit account.

Respectfully submitted,

Date: March 10, 2000

Stephen E. Reiter
Attorney for Applicants
Registration No. 31,192
Telephone: (858) 677-1409
Facsimile: (858) 677-1465

GRAY CARY WARE & FREIDENRICH LLP
4365 Executive Drive, Suite 1600
San Diego, CA 92121-2189

Gray Cary\GT\6172877.1
62574-157962

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL LABEL NUMBER **EL476992837US**

DATE OF DEPOSIT **March 10, 2000**

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1 10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT COMMISSIONER OF PATENTS, BOX PATENT APPLICATION, WASHINGTON, D C. 20231

LYNN MORKUNAS

(TYPED OR PRINTED NAME OF PERSON MAILING PAPER)

Lynn Morkunas

(SIGNATURE OF PERSON MAILING PAPER OR FEE)

APPLICATION

For

UNITED STATES LETTERS PATENT

on

A FAMILY OF TRANSCRIPTIONAL CO-REPRESSORS THAT INTERACT WITH NUCLEAR HORMONE RECEPTORS AND USES THEREFOR

by

Ronald M. Evans, J. Don Chen

and

Peter Ordentlich

Sheets of Drawings: Twelve (12)

Docket No.: SALK 1510-3

Attorneys

Gray Cary Ware & Freidenrich LLP
4365 Executive Drive, Suite 1600
San Diego, California 92121-2189

A Family of Transcriptional Co-repressors that Interact with Nuclear Hormone Receptors and Uses Therefor

RELATED APPLICATIONS

This application is a continuation-in-part application of pending United States application Serial No. 08/522,726, filed September 1, 1995 and is related to United States application Serial No. ______________, filed on even date herewith, each of which is incorporated herein in its entirety by reference.

FIELD OF THE INVENTION

The present invention relates to intracellular receptors, methods for the modulation thereof, and methods for the identification of novel ligands therefor. In a particular aspect, the present invention relates to methods for the identification of compounds which function as ligands (or ligand precursors) for intracellular receptors. In another aspect, the present invention relates to novel chimeric constructs and uses therefor.

BACKGROUND OF THE INVENTION

A central problem in eukaryotic molecular biology continues to be the elucidation of molecules and mechanisms that mediate specific gene regulation. As part of the scientific attack on this problem, a great deal of work has been done in efforts to identify ligands (i.e., exogenous inducers) which are capable of mediating specific gene regulation. Additional work has been done in efforts to identify other molecules involved in specific gene regulation.

Although much remains to be learned about the specifics of gene regulation, it is known that ligands modulate gene transcription by acting in concert with intracellular components, including intracellular receptors and discrete DNA sequences known as hormone response elements (HREs).

The identification of compounds that directly or indirectly interact with intracellular receptors, and thereby affect transcription of hormone-responsive genes, would be of significant value, e.g., for therapeutic applications.

Transcriptional silencing mediated by nuclear receptors plays an important role in development, cell differentiation, and is directly linked to the oncogenic activity of v-erbA. The mechanism underlying this effect is unknown but is one key to understanding the molecular basis of hormone action. Accordingly, the identification of components involved in transcriptional silencing would represent a great advance in current understanding of mechanisms that mediate specific gene regulation.

Other information helpful in the understanding and practice of the present invention can be found in commonly assigned United States Patent Nos. 5,071,773, 4,981,784, 5,260,432, and 5,091,513, all of which are hereby incorporated herein by reference in their entirety.

BRIEF DESCRIPTION OF THE INVENTION

The present invention overcomes many problems in the art by providing a family of receptor interacting co-repressors, referred to herein as "SMRT co-repressor", i.e., a silencing mediator (co-repressor) for retinoic acid receptor (RAR) and thyroid hormone receptor (TR). *In vivo*, members of the SMRT family of co-repressors function as potent co-repressors. A GAL4 DNA binding domain (DBD) fusion with a SMRT co-repressor behaves as a frank repressor of a GAL4-dependent reporter.

Together, these observations identify a novel family of cofactors that is believed to represent an important mediator of hormone action.

Accordingly, the present invention provides isolated silencing mediators of retinoic acid and thyroid hormone receptors, and isoforms or peptide portions thereof (SMRT co-repressors), that modulate transcriptional potential of members of the nuclear receptor superfamily. Such SMRT co-repressors comprise a repression domain having less than about 83% identity with a Sin3A interaction domain of N-CoR (amino acids 255 to 312 of SEQ ID NO: 11); less than about 57% identity with repression domain 1 of N-CoR (amino acids 1 to 312 of SEQ ID NO: 11); less than about 66% identity with a SANT domain of N-CoR (amino acids 312 to 668 of SEQ ID NO: 11) and/or; less than about 30% identity with repression domain 2 of N-CoR (amino acids 736 to 1031 of SEQ ID NO: 11).

In accordance with yet another embodiment of the present invention, there are provided isolated peptides comprising at least a portion of the invention SMRT co-repressor six contiguous amino acids of an amino acid sequence selected from the group consisting of:

amino acids 1 to 1030 of SEQ ID NO: 5;
amino acids 1 to 1029 of SEQ ID NO: 7;
amino acids 1 to 809 of SEQ ID NO: 9;
and conservative variations thereof,

provided the peptide is not identical to a sequence of SEQ ID NO: 11.

In addition, there are provided isolated antibodies that bind specifically to invention isolated peptides. There are also provided chimeric molecules comprising invention isolated peptides and at least a second molecule. Also provided are complexes comprising an invention SMRT co-repressor and a member of the superfamily of nuclear receptors and isolated antibodies that bind to such complexes.

Accordingly, the present invention provides isolated polynucleotides encoding members of the newly described family of silencing mediators of retinoic acid and thyroid hormone receptor or an isoform or peptide portion thereof (SMRT co-repressor), or an isolated polynucleotide complementary thereto. In addition, there are provided vectors comprising invention polynucleotides, as well as host cells containing invention polynucleotides.

In additional embodiments of the present invention, there are provided methods for identifying agents that modulate the repressor potential of a SMRT co-repressor.

In another embodiment according to the present invention, there are provided methods for identifying an agent that modulates a function of an invention SMRT co-repressor.

In another embodiment according to the present invention, there are provided methods of modulating the transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor) in a cell.

In another embodiment according to the present invention, there are provided methods of identifying a molecule that interacts specifically with a SMRT co-repressor.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the quantitation by phosphoimager of a dose-dependent dissociation of SMRT from RAR or TR by all-*trans* retinoic acid (atRA) or thyroid hormone (triiodothyronine or T3).

Figure 2 presents amino acid (aa) sequences of SMRT (Genbank accession number XXXXX). The aa sequence presented in parentheses (i.e., residues

1330-1376) is an alternatively spliced insert which is not present in the original two-hybrid clone (C-SMRT, aa 981 to C-terminal end). The proline-rich N-terminal domain (aa 1-160) and the glutamine-rich region (aa 1061-1132), as well as the ERDR and SG regions, are also indicated. The C-terminal region of SMRT (aa 1201 to C-terminal end) shows 48% aa identity to RIP13 (Seol et al., *Molecular Endocrinology* 9:72-85 (1995)). The rest of the sequence of RIP13 shows 22% aa identity to SMRT (aa 819-1200).

Figure 3 illustrates mediation of the silencing effect of hRARα and hTRβ by SMRT *in vivo*.

Figure 3(A) illustrates that v-erbA reverses the silencing effect of GAL-RAR (GAL4 DBD-hRARα 156-462) while SMRT restores the silencing effect.

Figure 3(B) illustrates that the RAR403 truncation mutant reverses the silencing effect of GAL-TR (GAL4 DBD-hTRβ 173-456) while SMRT restores the silencing effect.

Figure 3(C) illustrates that v-erbA and full length SMRT or C-SMRT have no effect on GAL-VP16 activity.

Figure 3(D) illustrates that a GAL4 DBD fusion of full length SMRT represses the thymidine kinase basal promoter activity containing four GAL4 binding sites. The fold of repression was calculated by dividing the normalized luciferase activity transfected with the GAL4 DBD alone by those transfected with indicated amount of GAL DBD fusion constructs.

Figure 4 provides an alignment of the human SMRT (SEQ ID NO: 5) and mouse SMRTα (SEQ ID NO: 7) amino acid sequences. Proteins were aligned using the CLUSTAL alignment program. Underlined sequence of mouse SMRTα corresponds to the amino acid sequences that are deleted in mouse SMRTβ. The

arrow indicates the start point of the previously described human SMRT co-repressor (sSMRT).

Figures 5A and 5B provide alignments of the human SMRT and human N-CoR co-repressors.

Figure 6A is a graph showing the results of transactivation experiments using transcripts encoding a detectable reporter and either wild type EcR (Ecr wt), a repression-Defective EcR allele Ecraa^{483T} (EcRA483T) or vp16 activation domain fused to Ultraspiracle (vp16-USP).

Figure 6B is a graph showing the results of transactivation experiments using CMV promoter-driven expression vectors. Wild-type EcR or EcR A483T was cotransfected with vp16-USP and Gal4-c-SMRT (aa 981 to C terminus) (Chen and Evans, *Nature* 377:454-457, (1995)) into CV-1 cells to examine its effect on the interaction with vertebrate corepressor. All cells were also cotransfected with a TK-luciferase reporter construct, pMH100-TK-Luc, containing four copies of the yeast Gal4-responsive element.

Figure 6C shows alignment of EcR, rTR, hRAR, and rRev-erbA receptor sequences and the secondary structure in the LBD signature motif region. Conserved residues are marked in dark. The mutation 483 (AT) is marked at the top of the corresponding residue.

Figure 7 is a graph showing β-galactosidase activity in a yeast two-hybrid screen with pAS-EcR as bait. pAS-EcR is a fusion gene with the region corresponding to aa 223-878 of EcRB1 fused C-terminally to the Gal4-DBD of the pAS1-CYH2 construct (Durfee et al., *Genes Dev* 7:555-569 (1993)); other Gal4-DBD-based nuclear receptor constructs used in this yeast two-hybrid assay include: USP (aa 50-508), hRAR (aa 186-462) and hTR (aa 121-410) (Schulman et al., *Proc. Natl. Acad. Sci. USA*, 92:8288-8292, (1995)), and SMRT (Chen and Evans, (1995), *supra*).

β-galactosidase activities were quantified by liquid assay for yeast cells treated either without ligand or with 3 μM of corresponding hormone. All-trans retinoic acid (ATRA) is a ligand of RAR; 3,3',5-triiodothyroacetic acid (T3) is a ligand of TR. RAR, retinoic acid receptor; TR, thyroid hormone receptor.

Figure 8A shows the complete amino acid sequence of the SMRTER protein (SEQ ID NO: 12). The underlined regions represent the residues also conserved in SMRT and N-CoR. The gray box indicates the sequences of the E52 clone.

Figure 8B is a schematic structural diagram of SMRTER, SMRT, and N-CoR showing the conserved SNOR, SANT, GST, ITS, D/ER repeat, and LSD motifs with their designated patterns positioned in their relative regions in each protein.

Figure 9. Sequence Comparison of SMRTER, SMRT, N-CoR, and Other Related Proteins. The SANT domains of various proteins are listed. Percent identities/similarities compared to SMRTER are shown on the right. Two potential helices are predicted in the N-terminal half of the SANT domain. Black boxes indicate identical sequences; gray boxes, similar or partially identical sequences.

Figure 10 is a schematic representation showing functional domains in SMRTER. Numbers on the left represent the regions in SMRTER used to generate the Gal4-DBD fusion genes. Black stippled bars indicate the locations of EcR-interacting domains; gray stippled bars indicate repression domains. Plus signs indicate that a positive interaction between SMRTER and the EcR complex and repression of basal activity by Gal4-SMRTER is significant. ERID = ecdysone receptor-interacting domain; SMRD = SMRTER repressor domain.

Figure 11A is a graph showing the interaction of ERID1 AND ERID2 with the EcR complex. Figure 11B is a graph showing the results of competition

between ERID1, ERID2 and c-SMRT for binding to EcR. Figure 11C is a graph showing that EcR A483T disrupts the interaction with ERID1 and ERID2.

Figure 12A shows the results of mapping three repression domains. To examine repressive activity, transcriptional activity of each Gal4-SMRTER fusion was compared to the basal activity of Gal4-DBD on reporter. Only repression with value approximately 5-fold or over is considered positive (+).

Figure 12B is a schematic representation of mapping the SMRTER-interacting domain in mSin3A and dSin3A. Yeast two-hybrid assays were used to assess the interaction between each Gal4-DBD-based fusion gene of each SMRD and the ACT-based fusion genes of mSin3A and dSin3A. The numbers indicate the region in either mSin3A or in dSin3A used to generate the ACT fusion genes. Constructs of mSin3A were described previously in Nagy et al., *Cell* 89:373-380, (1997).

Figure 12C shows an alignment of SMRD3 of SMRTER and an mSin3-interacting domain of N-CoR. Conserved residues are boxed in gray. An asterisk indicates the region where the mutation (Gly) was generated. Minus signs indicate that the interaction between SMRD3 and Sin3A was not detectable in the yeast two-hybrid assays. Repression was measured by comparing the transcriptional activity of Gal4-SMRD3 M2 or Gal4-SMRD3 M3 to that of wild-type Gal4-SMRD3 using transfection experiments as described above.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, there is provided a family of isolated SMRT co-repressors, and isoforms and peptide portions thereof, that modulate transcriptional potential of members of the nuclear receptor superfamily. Exemplary members of this family are co-repressors having substantially the same sequence as residues 1-1329 plus 1376-1495, as set forth in SEQ ID NO:1, optionally further

comprising the amino acid residues set forth in SEQ ID NO:2 (i. e., residues 1330-1375 of SEQ ID NO:1).

In another embodiment according to the present invention, the invention SMRT co-repressor comprises a repression domain having less than about 83% identity with a Sin3A interaction domain of N-CoR (as amino acids 255 to 312 of SEQ ID NO: 11); less than about 57% identity with repression domain 1 of N-CoR (amino acids 1 to 312 of SEQ ID NO: 11); less than about 66% identity with a SANT domain of N-CoR (amino acids 312 to 668 of SEQ ID NO: 11 and/or; less than about 30% identity with repression domain 2 of N-CoR (amino acids 736 to 1031 of SEQ ID NO: 11). Such an encoded SMRT co-repressor or peptide portion thereof is further characterized in that it can modulate transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor).

The invention SMRT co-repressors are additionally exemplified by a full length human SMRT co-repressor, (amino acids 1 to 2517 of SEQ ID NO: 5); and by two mouse SMRT isoforms, including a longer SMRT isoform designated mouse SMRTα, which has an amino acid sequence set forth as amino acids 1 to 2473 of SEQ ID NO: 7; and a shorter SMRT isoform designated mouse SMRTβ (amino acids 1 to 2253 of SEQ ID NO: 9). As compared to the mouse SMRTα isoform (SEQ ID NO: 7), the mouse SMRTβ isoform (SEQ ID NO: 9) has a deletion corresponding to amino acids 36 to 254 of SEQ ID NO: 7.

A peptide portion of a SMRT co-repressor is exemplified herein by amino acids 1 to 1031 of SEQ ID NO: 5; amino acids 1 to 1031 of SEQ ID NO: 7; and amino acids 1 to 813 of SEQ ID NO: 9, which includes the entire amino terminal domain of a SMRT co-repressor. Additional peptide portions of a SMRT co-repressor are exemplified by amino acids 1 to 303 of SEQ ID NO: 7; amino acids 845 to 986 of SEQ ID NO: 7; amino acids 427 to 663 of SEQ ID NO: 7; amino acids 845 to 1055 of SEQ ID NO: 7; amino acids 736 to 1031 of SEQ ID NO: 7; and amino acids 1 to 85 of SEQ ID NO: 9, which are sub-domains of the amino terminal domain

of mouse SMRTα that have nuclear receptor repressor potential, as well as by the corresponding peptide portions of human SMRT and corresponding peptide portions of mouse SMRTβ, which can modulate the transcriptional potential of a nuclear receptor, particularly a nuclear receptor that is in the form of a dimer, for example, a thyroid hormone receptor homodimer, a retinoic acid receptor homodimer, a retinoid X receptor homodimer, a thyroid hormone receptor-retinoid X receptor heterodimer, or a retinoic acid receptor-retinoid X receptor heterodimer. In addition, the invention relates to isolated peptides that contain at least six contiguous amino acids of an amino acid sequence set forth as amino acids 1 to 1030 of SEQ ID NO: 5; amino acids 1 to 1029 of SEQ ID NO: 5; or amino acids 1 to 809 of SEQ ID NO: 9, provided the SMRT peptide is not identical to a sequence of N-CoR (SEQ ID NO: 11).

Invention co-repressor can be an invertebrate SMRT co-repressor, such as the Drosophilia SMRTER co-repressor having an amino acid sequence as set forth in SEQ ID NO: 12, or conservative variations thereof.

Additional exemplary co-repressors are those containing one or both of the receptor interacting domains (ERID1 and ERID2) identified in the Drosophilia co-repressor. For example, co-repressors containing such receptor interacting domains can be selected from the following segments of the Drosophilia SMRTER co-repressor (SEQ. ID 12):

amino acids 1698-1924 of SEQ. ID NO:12,
amino acids 2951-3038 of SEQ. ID NO:12,
amino acids 1698-2063 of SEQ. ID NO:12,
amino acids 2094-3040 of SEQ. ID NO:12,
amino acids 2929-3181 of SEQ. ID NO:12,
amino acids 542-950 of SEQ. ID NO:12,
amino acids 2094-3181 of SEQ ID NO:12,
amino acids 2929-3040 of SEQ ID NO:12, and

amino acids 2951-3038 of SEQ ID NO:12,
and conservative variations thereof.

Additional exemplary co-repressors are those containing one or more of three autonomous repressor domains termed SMRD1, SMRD2, and SMRD3 identified in the SMRTER co-repressor. For example, invention co-repressors can contain the following autonomous repressor domains derived from Drosophilia SMRTER co-repressor (SEQ. ID 12):

amino acids 542-950 of SEQ. ID NO:12
amino acids 1698-1924 of SEQ ID NO:12,
amino acids 2951-3038 of SEQ. ID NO:12, and conservative variations thereof.

Conservative variations of the above-described SMRT co-repressors are also contemplated to be within the scope of the present invention. Moreover, proteins, polypeptides and peptides having at least 80% sequence identity with any of the SMRT co-repressors described herein are also contemplated to be within the scope of the invention.

In another embodiment according to the present invention, there are provided chimeric molecules comprising invention isolated peptides and at least a second molecule. For example, the second molecule in invention chimeric molecule can be a polynucleotide or a polypeptide. In one embodiment, the chimeric molecule is a fusion polypeptide comprising a SMRT co-repressor operably linked to a DNA binding domain of a transcription factor.

In another embodiment according to the present invention, there are provided isolated antibodies that bind specifically to invention isolated peptides. In one embodiment, an antibody of the invention binds specifically to an epitope of a SMRT co-repressor. Such an antibody is characterized, in part, in that it does not substantially crossreact with an N-CoR polypeptide. In another embodiment, an

antibody of the invention binds specifically to a complex, which includes a SMRT co-repressor or peptide portion thereof of the invention, a nuclear receptor and, optionally, a DNA regulatory element that is specifically bound by the nuclear receptor. Such an antibody is characterized, in part, in that it does not substantially crossreact with the nuclear receptor, either alone or bound to the DNA regulatory element. An antibody of the invention can be a monoclonal antibody, or can be one of a plurality of polyclonal antibodies, which essentially is a mixed population of monoclonal antibodies. The invention also relates to a cell line, which produces the monoclonal antibody of the invention.

Such antibodies can be employed for a variety of purposes, e.g., for studying tissue localization of invention SMRT co-repressor, the structure of functional domains, the purification of receptors, as well as in diagnostic applications, therapeutic applications, and the like. Preferably, for therapeutic applications, the antibodies employed will be monoclonal antibodies.

The above-described antibodies can be prepared employing standard techniques, as are well known to those of skill in the art, using the invention SMRT co-repressor or portions thereof as antigens for antibody production. Both anti-peptide and anti-fusion protein antibodies can be used [see, for example, Bahouth et al. (1991) Trends Pharmacol Sci. vol. 12:338-343; Current Protocols in Molecular Biology (Ausubel et al., eds.) John Wiley and Sons, New York (1989). Factors to consider in selecting portions of invention SMRT co-repressor for use as immunogen (as either a synthetic peptide or a recombinantly produced bacterial fusion protein) include antigenicity, accessibility (i.e., where the selected portion is derived from, e.g., the ligand binding domain, DNA binding domain, dimerization domain, and the like), uniqueness of the particular portion selected (relative to known receptors and co-repressors therefor), and the like.

In another embodiment according to the present invention, there are provided complexes comprising an invention SMRT co-repressor and a member of

the nuclear receptor superfamily and isolated antibodies that bind to such complexes. The nuclear receptor can be in the form of a monomer or dimer, for example, a thyroid hormone receptor homodimer, a retinoic acid receptor homodimer, a retinoid X receptor homodimer, a thyroid hormone receptor-retinoid X receptor heterodimer, a retinoic acid receptor-retinoid X receptor heterodimer, a ecdysone receptor-Ultraspiracle receptor heterodimer, and the like. Optionally or alternatively, the complex can include a DNA regulatory element, bound specifically by a DNA binding domain of the nuclear receptor.

The above-described complexes optionally further comprise a response element for the member of the nuclear receptor superfamily. Such response elements are well known in the art. Thus, for example, RAR response elements are composed of at least one direct repeat of two or more half sites separated by a spacer of five nucleotides. The spacer nucleotides can independently be selected from any one of A, C, G or T. Each half site of response elements contemplated for use in the practice of the invention comprises the sequence

-RGBNNM-,

wherein

R is selected from A or G;
B is selected from G, C, or T;
each N is independently selected from A, T, C, or G; and
M is selected from A or C;

with the proviso that at least 4 nucleotides of said -RGBNNM- sequence are identical with the nucleotides at corresponding positions of the sequence -AGGTCA-. Response elements employed in the practice of the present invention can optionally be preceded by N_x, wherein x falls in the range of 0 up to 5.

Similarly, TR response elements can be composed of the same half site repeats, with a spacer of four nucleotides. Alternatively, palindromic constructs as have been described in the art are also functional as TR response elements.

The above-described SMRT co-repressor/dimeric receptor complexes can be dissociated by contacting the complex with a ligand for the member of the nuclear receptor superfamily.

As employed herein, the term "ligand (or ligand precursor) for a member of the nuclear receptor superfamily" (i.e., intracellular receptor) refers to a substance or compound which, in its unmodified form (or after conversion to its "active" form), inside a cell, binds to receptor protein, thereby creating a ligand/receptor complex, which in turn can activate an appropriate hormone response element. A ligand therefore is a compound which acts to modulate gene transcription for a gene maintained under the control of a hormone response element, and includes compounds such as hormones, growth substances, non-hormone compounds that modulate growth, and the like. Ligands include steroid or steroid-like hormone, retinoids, thyroid hormones, pharmaceutically active compounds, and the like. Individual ligands may have the ability to bind to multiple receptors.

Accordingly, as employed herein, "putative ligand" (also referred to as "test compound") refers to compounds such as steroid or steroid-like hormones, pharmaceutically active compounds, and the like, that are suspected to have the ability to bind to the receptor of interest, and to modulate transcription of genes maintained under the control of response elements recognized by such receptor.

In another embodiment according to the present invention, there are provided polynucleotides encoding members of the above-described family of silencing mediators of retinoic acid and thyroid hormone receptor, or an isoform or peptide portion thereof (SMRT co-repressors), or an isolated polynucleotide complementary thereto.

Invention polynucleotides include those encoding a SMRT co-repressor comprises a repression domain having

a) less than about 83% identity with a Sin3A interaction domain of N-CoR set forth as amino acids 255 to 312 of SEQ ID NO: 11;

b) less than about 57% identity with repression domain 1 of N-CoR set forth as amino acids 1 to 312 of SEQ ID NO: 11;

c) less than about 66% identity with a SANT domain of N-CoR set forth as amino acids 312 to 668 of SEQ ID NO: 11; or

d) less than about 30% identity with repression domain 2 of N-CoR set forth as amino acids 736 to 1031 of SEQ ID NO: 11.

In addition, an invention polynucleotide can encode a mouse SMRTβ isoform having an amino acid sequence as set forth in SEQ ID NO: 9 or conservative variations thereof, or a polynucleotide having a nucleotide sequence as set forth in SEQ ID NO: 8.

Further examples of invention polynucleotides are those comprising a nucleotide sequence selected from the group consisting of:

nucleotides 1 to 3094 of SEQ ID NO: 4;
nucleotides 1 to 3718 of SEQ ID NO: 6;
nucleotides 1 to 2801 of SEQ ID NO: 8;
nucleotides 1 to 8388 of SEQ ID NO: 6;
nucleotides 1 to 7465 of SEQ ID NO: 8; and
nucleotides 1 to 8561 of SEQ ID NO: 4.

The invention polynucleotides further comprise those encoding a human SMRT co-repressor having an amino acid sequence as set forth in SEQ ID NO: 5, for example, a nucleotide sequence as set forth in SEQ ID NO: 4; by a polynucleotide encoding a mouse SMRTα isoform having an amino acid sequence as set forth in SEQ ID NO: 7, for example, a nucleotide sequence as set forth in SEQ ID NO: 6; and by a polynucleotide encoding a mouse SMRTβ isoform having an amino acid sequence as set forth in SEQ ID NO: 9, for example, a nucleotide sequence as set forth in SEQ ID NO: 8. A polynucleotide of the invention is further exemplified by

polynucleotides encoding peptide portions of a SMRT co-repressor such as a polynucleotide containing nucleotides 1 to 3094 of SEQ ID NO: 4; nucleotides 1 to 3718 of SEQ ID NO: 7; or nucleotides 1 to 2801 of SEQ ID NO: 8, which can repress the transcriptional activity of nuclear receptor, particularly a nuclear receptor that is in the form of dimer.

Additional invention polynucleotides include those encoding a full length insect SMRTER co-repressor having an amino acid sequence as set forth in SEQ ID NO: 12, or conservative variations thereof.

Additional exemplary invention polynucleotides are those encoding one or both of the receptor interacting domains (ERID1 and ERID2) identified in invention co-repressors. For example, polynucleotides encoding such receptor interacting domains can be selected from those encoding the following segments of the Drosophilia SMRTER co-repressor (SEQ. ID 12):

amino acids 1698-1924 of SEQ. ID NO:12,
amino acids 2951-3038 of SEQ. ID NO:12,
amino acids 1698-2063 of SEQ. ID NO:12,
amino acids 2094-3040 of SEQ. ID NO:12,
amino acids 2929-3181 of SEQ. ID NO:12,
amino acids 542-950 of SEQ. ID NO:12,
amino acids 2094-3181 of SEQ ID NO:12,
amino acids 2929-3040 of SEQ ID NO:12, and
amino acids 2951-3038 of SEQ ID NO:12,
and conservative variations thereof.

Additional exemplary invention polynucleotides are those encoding one or more of three autonomous repressor domains termed SMRD1, SMRD2, and SMRD3 identified in the invention co-repressors. For example, polynucleotides encoding such autonomous repressor domains can be selected from those encoding the following segments of the Drosophilia SMRTER co-repressor (SEQ. ID 12):

amino acids 542-950 of SEQ. ID NO:12
amino acids 1698-1924 of SEQ ID NO:12,
amino acids 2951-3038 of SEQ. ID NO:12, and conservative variations thereof.

A polynucleotide that has at least 80% sequence identity or that hybridizes, (preferably under high stringency conditions) with any one of the above-described polynucleotides is also contemplated to be within the scope of this invention.

A polynucleotide of the invention can be operably linked to a second nucleotide sequence and, therefore, can encode a fusion polypeptide, for example, a SMRT co-repressor, or peptide portion thereof, operably linked to a DNA binding domain of a transcription factor.

Additional examples of invention isolated oligonucleotides, are those which generally are at least about 15 nucleotides in length and can hybridize specifically to the polynucleotide of the invention, but not to a polynucleotide encoding an N-CoR polypeptide (SEQ ID NO: 11). An oligonucleotide of the invention can be useful as a probe, or as a primer for a PCR procedure, or can encode a peptide containing at least five contiguous amino acids of a SMRT co-repressor. In one embodiment, an oligonucleotide of the invention encodes at least five contiguous amino acids of a sequence such as that shown as amino acids 720 to 745 of SEQ ID NO: 5; or amino acids 716 to 742 of SEQ ID NO: 7; or amino acids 497 to 523 of SEQ ID NO: 9. In another embodiment, an oligonucleotide of the invention can hybridize specifically to a polynucleotide encoding human SMRT (SEQ ID NO: 5) or mouse SMRTα (SEQ ID NO: 7), and, optionally, to a polynucleotide encoding mouse SMRTβ (SEQ ID NO: 9).

The phrase "substantially the same" as used herein in reference to a nucleotide sequence of DNA, a ribonucleotide sequence of RNA, or an amino acid

sequence of protein, means sequences that have slight and non-consequential sequence variations from the actual sequences disclosed herein. Species that are substantially the same are considered to be equivalent to the disclosed sequences and as such are within the scope of the appended claims. In this regard, "slight and non-consequential sequence variations" means that sequences substantially the same as the DNA, RNA, or proteins disclosed and claimed herein are functionally equivalent to the sequences disclosed and claimed herein. Functionally equivalent sequences will function in substantially the same manner to produce substantially the same compositions as the nucleic acid and amino acid compositions disclosed and claimed herein. In particular, functionally equivalent DNAs encode proteins that are the same as those disclosed herein or that have conservative amino acid variations, such as substitution of a non-polar residue for another non-polar residue or a charged residue for a similarly charged residue. These changes include those recognized by those of skill in the art as those that do not substantially alter the tertiary structure of the protein.

In another embodiment according to the present invention, there are provided vectors comprising an invention polynucleotide, and host cells containing invention polynucleotides. The invention vector can be an expression vector, including, for example, a viral vector, and the polynucleotide, or a vector containing the polynucleotide, can be contained in a host cell. In one embodiment, the polynucleotide of the invention is operably linked to a tissue specific DNA regulatory element. In another embodiment, a SMRT co-repressor or peptide portion thereof encoded by the polynucleotide is expressed in a host cell.

In another embodiment according to the present invention, there are provided methods for identifying an agent that modulates the repressor potential of a SMRT co-repressor. In this embodiment, the invention method comprises contacting a host cell with an agent, and detecting a change in the level of expression of a first expressible nucleotide sequence in response to the agent, thereby identifying an agent that modulates the repressor potential of a SMRT co-repressor. In such a method, the host cell is characterized, in part, in that it contains a first expressible nucleotide

sequence operably linked to a first DNA regulatory element, and expresses a fusion polypeptide composed of an invention SMRT co-repressor, or peptide portion thereof, and a DNA binding domain of a first transcription factor that can specifically bind the first DNA regulatory element. Binding of the DNA binding domain of the first transcription factor to the first DNA regulatory element results in expression of the first expressible nucleotide sequence in the host cell.

In another embodiment according to the present invention, there are provided methods for identifying an agent that modulates a function of an invention SMRT co-repressor. In this embodiment, the invention method comprises contacting an invention SMRT co-repressor, a member of the nuclear receptor superfamily, and an agent, and detecting an altered activity of the SMRT co-repressor in the presence of the agent as compared to the absence of the agent, thereby identifying an agent that modulates a function of the SMRT co-repressor.

A method of the invention can be performed, for example, by contacting a host cell with an agent, and detecting a change in the level of expression of a first expressible nucleotide sequence in response to the agent, thereby identifying an agent that modulates the repressor potential of a SMRT co-repressor. In such a method, the host cell is characterized, in part, in that it contains a first expressible nucleotide sequence operably linked to a first DNA regulatory element, and expresses a fusion polypeptide composed of a SMRT co-repressor or peptide portion thereof of the invention, and a DNA binding domain of a first transcription factor, which can specifically bind the first DNA regulatory element; binding of the DNA binding domain of the first transcription factor to the first DNA regulatory element results in expression of the first expressible nucleotide sequence in the host cell. The first expressible nucleotide sequence can be an endogenous gene, which is normally present in the host cell, or can be a sequence that has been introduced into the host cell, either transiently or stably, using methods of recombinant DNA technology. In one embodiment, the first DNA binding domain is a GAL4 DNA binding domain and the first DNA regulatory element is a GAL4 DNA regulatory element that is operably

linked to an expressible nucleotide sequence, for example, a reporter gene, and is introduced into the host cell.

Thus, the invention method can identify an agent that increases or decreases the repressor potential of the SMRT co-repressor, or of an agent that increases or decreases the function of the SMRT co-repressor. The agent can directly interact with the SMRT co-repressor or peptide portion thereof, thereby modulating the repressor potential or function of the SMRT co-repressor, or can interact with a cellular molecule that, in turn, can alter the repressor potential or function of a SMRT co-repressor, thereby increasing or decreasing the repressor potential of the SMRT co-repressor.

The host cell can optionally contain a second expressible nucleotide sequence operably linked to a second DNA regulatory element, and can express a second fusion polypeptide, which is composed of an N-CoR polypeptide, or a repressor domain thereof, and a DNA binding domain of a second transcription factor, which can specifically bind the second DNA regulatory element. By comparing the level of expression of the first expressible nucleotide sequence and the second expressible nucleotide sequence in the host cell upon contacting the host cell with the agent, an agent that independently or coordinately modulates SMRT and N-CoR repressor activity. For example, detecting a change in the level of expression of the first expressible nucleotide sequence, but not in the level of expression of the second expressible nucleotide sequence, due to contacting the host cell with the agent identifies an agent that modulates the repressor potential of a SMRT co-repressor, but not of an N-CoR polypeptide can be identified.

In practicing a method of the invention, the SMRT co-repressor, or peptide portion thereof, can be, for example, an amino acid sequence such as amino acids 1 to 1031 of SEQ ID NO: 5; amino acids 1 to 1031 of SEQ ID NO: 7; or amino acids 1 to 813 of SEQ ID NO: 9. The agent can be, for example, an antibody or antigen binding fragment thereof, a peptide, or a small organic molecule.

In another embodiment according to the present invention, there are provided methods of modulating the transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor) in a cell, the method comprising introducing an invention isolated polynucleotide into the cell, whereby the polynucleotide or an expression product of the polynucleotide alters the level of a SMRT co-repressor in the cell, thereby modulating the transcriptional potential of the nuclear receptor.

In another embodiment according to the present invention, there are provided methods of modulating the transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor) in a cell, the method comprising introducing an invention isolated polynucleotide into the cell, whereby the polynucleotide or an expression product of the polynucleotide alters the level of a SMRT co-repressor in the cell, thereby modulating the transcriptional potential of the nuclear receptor.

In performing a method of the invention, an agent that alters an interaction of the SMRT co-repressor, or peptide portion thereof, with the nuclear receptor can be identified using a binding assay, such as an electrophoretic mobility shift assay wherein the level of expression of an expressible nucleotide sequence. Such a method can also identify an agent that alters the ability of the invention SMRT co-repressor, or peptide portion thereof, to interact specifically with the nuclear receptor, but does not alter the level of expression of the expressible nucleotide sequence; or an agent that alters the level of expression of the expressible nucleotide sequence, but does not alter interaction of the SMRT co-repressor or peptide portion thereof with the nuclear receptor; or an agent that alters an interaction of the SMRT co-repressor, or peptide portion thereof, with the nuclear receptor and alters the level of expression of the expressible nucleotide sequence. The agent can, but need not be, a ligand for the nuclear receptor, and the method can be performed in a cell or in a reaction mixture *in vitro*.

Alternatively, an invention polynucleotide can be introduced into the cell, whereby the polynucleotide, or an expression product of the polynucleotide, alters the level of a SMRT co-repressor in the cell, thereby modulating the transcriptional potential of the nuclear receptor. The polynucleotide can encode an invention SMRT co-repressor or peptide, portion thereof, which can be expressed in the cell, thereby increasing the level of a SMRT co-repressor, or peptide portion thereof, in the cell. The polynucleotide also can be an antisense polynucleotide, that decreases the level of a SMRT co-repressor in the cell.

In another embodiment according to the present invention, there are provided methods of identifying a molecule that interacts specifically with a SMRT co-repressor. In this embodiment, invention methods comprise contacting the molecule with an invention SMRT co-repressor and detecting specific binding of the molecule to the SMRT co-repressor, thereby identifying a molecule that interacts specifically with a SMRT co-repressor.

The molecule can be any molecule that interacts specifically with a SMRT co-repressor, including, for example, a small organic molecule such as a drug, a peptide, a nucleic acid molecule, and the like. In one embodiment, the molecule is a cellular factor, for example, a cellular protein that modulates the ability of a SMRT co-repressor to repress transcriptional activity of a nuclear receptor. In another embodiment, the method further involves isolating the molecule that interacts specifically with the SMRT co-repressor or peptide portion thereof.

In accordance with yet another aspect of the present invention, there are provided methods to block the repressing effect of invention SMRT co-repressors, said method comprising administering an effective amount of an antibody as described herein. Alternatively, a silencing domain of a nuclear receptor can be employed. Those of skill in the art can readily determine suitable methods for administering said antibodies, and suitable quantities for administration, which will vary depending on

numerous factors, such as the indication being treated, the condition of the subject, and the like.

In accordance with another aspect of the present invention, there is provided a method to repress (or silence) the activity of a member of the nuclear receptor superfamily containing a silencing domain that represses basal level promoter activity of target genes, said method comprising contacting said member of the nuclear receptor superfamily with a sufficient quantity of an invention SMRT co-repressor so as to repress the activity of said member. Members of the nuclear receptor superfamily contemplated for repression in accordance with this aspect of the present invention include, for example, thyroid hormone receptor, retinoic acid receptor, vitamin D receptor, peroxisome proliferator activated receptor, and the like.

In accordance with yet another aspect of the present invention, there is provided a method to identify compounds which relieve the repression of nuclear receptor activity caused by an invention SMRT co-repressor, said method comprising comparing the size of the SMRT co-repressor/dimeric receptor complex (i.e., complexes comprising the invention SMRT co-repressor and a homodimeric or heterodimeric member of the nuclear receptor superfamily) upon exposure to test compound, relative to the size of said complex in the absence of test compound. An observed size corresponding to intact complex is indicative of an inactive compound, while an observed size that reflects dissociation of the complex is indicative of a compound that disrupts the complex, thereby relieving the repression caused thereby. Optionally, the complex employed in this assay further comprises a response element for said member of the nuclear receptor superfamily.

The size of the above-described complex can readily be determined employing various techniques available in the art. For example, electrophoretic mobility shift assays (EMSA) can be employed (wherein receptor alone or receptor-SMRT co-repressor complex is bound to target DNA and the relative mobility thereof determined).

Those of skill in the art can readily identify other methodology which can be employed to determine the size of the complex as a result of exposure to putative ligand.

In accordance with a still further aspect of the present invention, there is provided a method to identify compounds which relieve the repression of nuclear receptor activity caused by an invention SMRT co-repressor, without substantially activating said receptor, said method comprising:

comparing the reporter signal produced by two different expression systems in the absence and presence of test compound,

wherein said first expression system comprises a complex comprising:

a homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer, or retinoic acid receptor-retinoid X receptor heterodimer,

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a reporter gene, and

optionally, invention SMRT co-repressor, and

wherein said second expression system comprises a complex comprising:

a homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first expression system, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes,

the same response element-reporter combination as employed in said first expression system, and

optionally, invention SMRT co-repressor, and thereafter selecting those compounds which provide:

a higher reporter signal upon exposure of said compound to said first expression system, relative to reporter signal in the absence of said compound, and

substantially the same reporter signal upon exposure of said compound to said second expression system, relative to reporter signal in the absence of said compound,

wherein said selected compounds are capable of relieving the repression of nuclear receptor activity caused by a SMRT co-repressor having a structure and function characteristic of an invention SMRT co-suppressor but substantially lacking the ability to activate nuclear receptor activity.

The addition of invention SMRT co-repressor is optional in the above-described assay because it is present endogenously in most host cells employed for such assays. It is preferred, to ensure the presence of a fairly constant amount of SMRT co-repressor, and to ensure that SMRT co-repressor is not a limiting reagent, that SMRT co-repressor be supplied exogenously to the above-described assays.

Mutant receptors contemplated for use in the practice of the present invention are conveniently produced by expression plasmids, introduced into the host cell by transfection. Mutant receptors contemplated for use herein include RAR403 homodimers, RAR403-containing heterodimers, TR160 homodimers, TR160-containing heterodimers, and the like.

Reporter constructs contemplated for use in the practice of the present invention comprise:

(a) a promoter that is operable in the host cell,

(b) a hormone response element, and

(c) a DNA segment encoding a reporter protein,

wherein the reporter protein-encoding DNA segment is operatively linked to the promoter for transcription of the DNA segment, and

wherein the hormone response element is operatively linked to the promoter for activation thereof.

Hormone response elements contemplated for use in the practice of the present invention are well known in the art, as has been noted previously.

Exemplary reporter genes include chloramphenicol transferase (CAT), luciferase (LUC), beta-galactosidase (β -gal), and the like. Exemplary promoters include the simian virus (SV) promoter or modified form thereof (e.g., SV), the thymidine kinase (TK) promoter, the mammary tumor virus (MTV) promoter or modified form thereof (e.g., ΔMTV), and the like [see, for example, Mangelsdorf et al., in Nature 345:224-229 (1990), Mangelsdorf et al., in Cell 66:555-561 (1991), and Berger et al., in J. Steroid Biochem. Molec. Biol. 41:733-738 (1992).

As used herein in the phrase "operative response element" or "operatively linked" the word "operative" means that the respective DNA sequences (represented by the terms "GAL4 response element" and "reporter gene") are operational, i.e., work for their intended purposes; such that after the two segments are linked, upon appropriate activation by a ligand-receptor complex, the reporter gene will be expressed as the result of the fact that the "GAL4 response element" was "turned on" or otherwise activated.

In practicing the above-described functional bioassay, the expression plasmid and the reporter plasmid are co-transfected into suitable host cells. The transfected host cells are then cultured in the presence and absence of a test compound to determine if the test compound is able to produce activation of the promoter operatively linked to the response element of the reporter plasmid. Thereafter, the transfected and

cultured host cells are monitored for induction (i.e., the presence) of the product of the reporter gene sequence.

Any cell line can be used as a suitable "host" for the functional bioassay contemplated for use in the practice of the present invention. Thus, cells contemplated for use in the practice of the present invention include transformed cells, non-transformed cells, neoplastic cells, primary cultures of different cell types, and the like. Exemplary cells which can be employed in the practice of the present invention include Schneider cells, CV-1 cells, HuTu80 cells, F9 cells, NTERA2 cells, NB4 cells, HL-60 cells, 293 cells, Hela cells, yeast cells, and the like. Preferred host cells for use in the functional bioassay system are COS cells and CV-1 cells. COS-1 (referred to as COS) cells are monkey kidney cells that express SV40 T antigen (Tag); while CV-1 cells do not express SV40 Tag. The presence of Tag in the COS-1 derivative lines allows the introduced expression plasmid to replicate and provides a relative increase in the amount of receptor produced during the assay period. CV-1 cells are presently preferred because they are particularly convenient for gene transfer studies and provide a sensitive and well-described host cell system.

The above-described cells (or fractions thereof) are maintained under physiological conditions when contacted with physiologically active compound. "Physiological conditions" are readily understood by those of skill in the art to comprise an isotonic, aqueous nutrient medium at a temperature of about 37°C.

In accordance with yet another aspect of the present invention, there is provided a method to identify compounds which activate nuclear receptor activity, but substantially lack the ability to relieve the repression caused by an invention SMRT co-repressor, said method comprising:

comparing the reporter signal produced by two different expression systems in the absence and presence of test compound,

wherein said first expression system comprises a complex comprising:

a homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer, or retinoic acid receptor-retinoid X receptor heterodimer,

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a reporter, and

optionally, invention SMRT co-repressor, and

wherein said second expression system comprises a complex comprising:

a homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first expression system, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes,

the same response element-reporter combination as employed in said first expression system, and

optionally, invention SMRT co-repressor, and thereafter

selecting those compounds which provide:

a higher reporter signal upon exposure of said compound to said second expression system, relative to reporter signal in the absence of compound, and

substantially the same reporter signal upon exposure of said compound to said first expression system, relative to reporter signal in the absence of said compound,

wherein said selected compounds are capable of activating nuclear receptor activity, but substantially lacking the ability to relieve the repression caused by a SMRT co-repressor having a structure and function characteristic of, an invention SMRT co-repressor for retinoic acid and thyroid receptors.

In accordance with a still further aspect of the present invention, there is provided a method to identify compounds which relieve the repression of nuclear receptor activity caused by an invention SMRT co-repressor, and activate said receptor, said method comprising:

comparing the reporter signal produced by two different expression systems in the absence and presence of test compound,

wherein said first expression system comprises a complex comprising:

a homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer, or retinoic acid receptor-retinoid X receptor heterodimer,

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a reporter, and

optionally, invention SMRT co-repressor, and

wherein said second expression system comprises a complex comprising:

a homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first expression system, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes,

the same response element-reporter combination as employed in said first expression system, and

optionally, invention SMRT co-repressor, and thereafter

selecting those compounds which provide:

increased reporter signal upon exposure of said compound to said second expression system, relative to reporter signal in the absence of said compound, and

substantially increased reporter signal upon exposure of said compound to said first expression system, relative to reporter signal in the absence of said compound,

wherein said selected compounds are capable of relieving the repression of nuclear receptor activity caused by a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors, and activating said receptor.

In accordance with still another embodiment of the present invention, there are provided modified forms of the above-described SMRT co-repressor, including:

full length silencing mediator for retinoic acid and thyroid receptors plus GAL4 DNA binding domain,

full length silencing mediator for retinoic acid and thyroid receptors plus GAL4 activation domain,

full length silencing mediator for retinoic acid and thyroid receptors plus glutathione S-transferase (GST) tag,

and the like.

The above-described modified forms of invention SMRT co-repressor can be used in a variety of ways, e.g., in the assays described herein.

An especially preferred modified SMRT co-repressor of the invention comprises full length silencing mediator for retinoic acid and thyroid receptors plus GAL4 activation domain.

In accordance with a still further embodiment of the present invention, there is provided a method to identify compounds which disrupt the ability of an invention SMRT co-repressor to complex with nuclear receptors, without substantially activating said receptor, said method comprising:

comparing the reporter signal produced by two different expression systems in the absence and presence of test compound,

wherein said first expression system comprises a complex comprising:

a modified SMRT co-repressor as described above,

a homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer or retinoic acid receptor-retinoid X receptor heterodimer, and

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a reporter, and

wherein said second expression system comprises a complex comprising:

said modified SMRT co-repressor,

a homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first expression system, wherein said member is mutated such that it retains hormone dependent activation activity but has lost

its ability to repress basal level promoter activity of target genes, and

the same response element-reporter combination as employed in said first expression system, and thereafter

selecting those compounds which provide:

a lower reporter signal upon exposure of said compound to said first expression system, relative to reporter signal in the absence of said compound, and

substantially the same reporter signal upon exposure of said compound to said second expression system, relative to reporter signal in the absence of said compound,

wherein said selected compounds are capable of disrupting the ability of a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors to complex with nuclear receptors, without substantially activating said receptor.

Mutant receptors contemplated for use in this embodiment of the present invention include RAR403 homodimers, RAR403-containing heterodimers, TR160 homodimers, TR160-containing heterodimers, and the like.

Suitable host cells for use in this embodiment of the present invention include mammalian cells as well as yeast cells. Yeast cells are presently preferred because they introduce no background since SMRT (i.e., silencing mediator (SMRT co-repressor) for retinoic acid receptor (RAR) and thyroid hormone receptor (TR)) is not endogenous to yeast.

In accordance with yet another embodiment of the present invention, there is provided a method to identify compounds which activate nuclear receptor

activity, but substantially lack the ability to disrupt a complex comprising a nuclear receptor and an invention SMRT co-repressor, said method comprising:

comparing the reporter signal produced by two different expression systems in the absence and presence of test compound,

wherein said first expression system comprises a complex comprising:

a modified SMRT co-repressor as described above,

a homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer or retinoic acid receptor-retinoid X receptor heterodimer, and

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a reporter, and

wherein said second expression system comprises:

said modified SMRT co-repressor,

a homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first expression system, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes, and

the same response element-reporter combination as employed in said first expression system, and thereafter

selecting those compounds which provide:

a higher reporter signal upon exposure of said compound to said second expression system, relative to reporter signal in the absence of compound, and

substantially the same reporter signal upon exposure of said compound to said first expression system, relative to reporter signal in the absence of compound,

wherein said selected compounds are capable of activating nuclear receptor activity, but substantially lack the ability to disrupt the complex of an invention SMRT co-repressor.

Suitable host cells for use in this embodiment of the present invention include mammalian cells as well as yeast cells. Yeast cells are presently preferred because they introduce no background since SMRT is not endogenous to yeast.

In accordance with a still further embodiment of the present invention, there is provided a method to identify compounds which activate a nuclear receptor, and disrupt the ability of an invention SMRT co-repressor to complex with said receptor, said method comprising:

comparing the reporter signal produced by two different expression systems in the absence and presence of test compound,

wherein said first expression system comprises a complex comprising:

a modified SMRT co-repressor as described above,

a homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer or retinoic acid receptor-retinoid X receptor heterodimer, and

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a reporter, and

wherein said second expression system comprises a complex comprising:

said modified SMRT co-repressor,

the same homodimeric or heterodimeric member of the nuclear receptor superfamily as employed in said first expression system, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes, and

the same response element-reporter combination as employed in said first expression system, and thereafter

selecting those compounds which provide:

a reduction in reporter signal upon exposure of compound to said first expression system, relative to reporter signal in the absence of said compound, and

increased reporter signal upon exposure of compound to said second expression system, relative to reporter signal in the absence of said compound,

wherein said selected compounds are capable of activating a nuclear receptor and disrupting a complex comprising nuclear receptor and a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors.

Suitable host cells for use in this embodiment of the present invention include mammalian cells as well as yeast cells. Yeast cells are presently preferred because they introduce no background since SMRT is not endogenous to yeast.

In accordance with yet another aspect of the present invention, there is provided a method to identify compounds which activate a nuclear receptor and/or disrupt the ability of an invention SMRT co-repressor to complex with said receptor, said method comprising:

comparing the reporter signals produced by a combination expression system in the absence and presence of test compound,

wherein said combination expression system comprises:

a first homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer, or retinoic acid receptor-retinoid X receptor heterodimer,

a second homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first homodimer or heterodimer, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes (i.e., provides basal level expression),

wherein either said first homodimer (or heterodimer) or said second homodimer (or heterodimer) is operatively linked to a GAL4 DNA binding domain,

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a first reporter,

a GAL4 response element, wherein said response element is operatively linked to a second reporter, and

optionally a SMRT co-repressor of nuclear receptor activity, said SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors, and thereafter

identifying as capable of relieving the repression of nuclear receptor activity caused by a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors, but substantially lacking the ability to activate nuclear receptor activity those compounds which provide:

a higher reporter signal from the reporter responsive to the first member upon exposure of said compound to said first member, relative to reporter signal in the absence of said compound, and

substantially the same reporter signal from the reporter responsive to the second member upon exposure of said compound to said second member, relative to reporter signal in the absence of said compound, or

identifying as capable of activating nuclear receptor activity, but substantially lacking the ability to relieve the repression caused by a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors those compounds which provide:

a higher reporter signal from the reporter responsive to the second member upon exposure of said compound to said second member, relative to reporter signal in the absence of compound, and

substantially the same reporter signal from the reporter responsive to the first member upon exposure of said compound to said first member, relative to reporter signal in the absence of said compound, or

identifying as capable of relieving the repression of nuclear receptor activity caused by a SMRT co-repressor having a structure and function characteristic of

the silencing mediator for retinoic acid and thyroid receptors, and activating said receptor those compounds which provide:

a higher reporter signal from the reporter responsive to the second member upon exposure of said compound to said second member, relative to reporter signal in the absence of said compound, and

a greater increase in reporter signal from the reporter responsive to the first member upon exposure of said compound to said first member, relative to reporter signal in the absence of said compound.

Thus, the change in expression level of the two different reporters introduced in a single transfection can be monitored simultaneously. Based on the results of this single transfection, one can readily identify the mode of interaction of test compound with the receptor/SMRT complex.

Exemplary GAL4 response elements are those containing the palindromic 17-mer:

5'-CGGAGGACTGTCCTCCG-3' (SEQ ID NO:3),

such as, for example, 17MX, as described by Webster et al., in *Cell* **52**:169-178 (1988), as well as derivatives thereof. Additional examples of suitable response elements include those described by Hollenberg and Evans in *Cell* **55**:899-906 (1988); or Webster et al. in *Cell* **54**:199-207 (1988).

In accordance with still another embodiment of the present invention, there is provided a method to identify compounds which activate a nuclear receptor and/or disrupt the ability of an invention SMRT co-repressor to complex with said receptor, said method comprising:

comparing the reporter signals produced by a combination expression system in the absence and presence of test compound,

wherein said combination expression system comprises:

a modified SMRT co-repressor as described above,

a first homodimeric or heterodimeric member of the nuclear receptor superfamily selected from thyroid hormone receptor homodimer, thyroid hormone receptor-retinoid X receptor heterodimer, retinoic acid receptor homodimer, or retinoic acid receptor-retinoid X receptor heterodimer,

a second homodimeric or heterodimeric form of the same member of the nuclear receptor superfamily as employed in said first homodimer or heterodimer, wherein said member is mutated such that it retains hormone dependent activation activity but has lost its ability to repress basal level promoter activity of target genes,

wherein either said first homodimer (or heterodimer) or said second homodimer (or heterodimer) is operatively linked to a GAL4 DNA binding domain,

a response element for said member of the nuclear receptor superfamily, wherein said response element is operatively linked to a first reporter,

a GAL4 response element, wherein said response element is operatively linked to a second reporter, and thereafter

identifying as capable of disrupting the ability of a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors to complex with a nuclear receptor, without substantially activating nuclear receptor, those compounds which provide:

a lower reporter signal from the reporter responsive to the first member upon exposure of said compound to said first member, relative to reporter signal in the absence of said compound, and

substantially the same reporter signal from the reporter responsive to the second member upon exposure of said compound to said second member, relative to reporter signal in the absence of said compound, or

identifying as capable of activating nuclear receptor activity, but substantially lacking the ability to disrupt a complex comprising a nuclear receptor and a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors, those compounds which provide:

a higher reporter signal from the reporter responsive to the second member upon exposure of said compound to said second member, relative to reporter signal in the absence of compound, and

substantially the same reporter signal from the reporter responsive to the first member upon exposure of said compound to said first member, relative to reporter signal in the absence of said compound, or

identifying as capable of disrupting a complex comprising a nuclear receptor and a SMRT co-repressor having a structure and function characteristic of the silencing mediator for retinoic acid and thyroid receptors, and activating said receptor those compounds which provide:

a reduction in reporter signal from the reporter responsive to the first member upon exposure of said compound to said first member, relative to reporter signal in the absence of said compound, and

increased reporter signal from the reporter responsive to the second member upon exposure of said compound to said second member, relative to reporter signal in the absence of said compound.

In accordance with a still further aspect of the present invention, there is provided a method to identify compounds which relieve the repression of nuclear

receptor activity caused by an invention SMRT co-repressor, said method comprising determining the effect of adding test compound to an expression system comprising:

a modified member of the nuclear receptor superfamily, wherein said modified member contains an activation domain which renders said receptor constitutively active,

a fusion protein comprising the receptor interaction domain of SMRT operatively linked to the GAL4 DNA binding domain, and

a GAL4 response element operatively linked to a reporter.

Prior to addition of an effective ligand for the member of the nuclear receptor superfamily employed herein, the association of the modified member and the fusion protein will be effective to bind the GAL4 response element and activate transcription of the reporter. The presence of an effective ligand is indicated by a reduction of reporter signal upon exposure to ligand, which disrupts the interaction of the modified member and fusion protein.

Activation domains contemplated for use in the practice of the present invention are well known in the art and can readily be identified by the artisan. Examples include the GAL4 activation domain, BP64, and the like.

To summarize, a novel family of nuclear receptor SMRT co-repressor which mediates the transcriptional silencing of RAR and TR has been identified. This discovery is of great interest because transcriptional silencing has been shown to play an important role in development, cell differentiation and the oncogenic activity of v-erbA (Baniahmad et al., *EMBO J.* **11**:1015-1023 (1992)); Gandrillon et al., *Cell* **49**:687-697 (1989)); Zenke et al., *Cell* **61**:1035-1049 (1990); Barlow et al., *EMBO J.* **13**:4241-4250 (1994); Levine and Manley, *Cell* **59**:405-408 (1989); Baniahmad et al., *Proc. Natl. Acad. Sci. USA* **89**:10633-10637 (1992b); and Saitou et al., *Nature* **374**:159-162 (1995)). In fact, v-erbA mutants that harbor the Pro160->Arg change in the TR neither repress basal

transcription nor are capable of oncogenic transformation (Damm and Evans, (1993), *supra*).

The function of SMRT as a silencing mediator (co-repressor) of RAR and TR is analogous to mSin3 in the Mad-Max-Sin3 ternary complex (Schreiber-Agus et al., *Cell* **80**:777-786 (1995); and Ayer et al., *Cell* **80**:767-776 (1995)). Because GAL-SMRT functions as a potent repressor when bound to DNA, it is reasonable to speculate that the function of the unliganded receptors is to bring with them SMRT to the template via protein-protein interaction. Thus, the repressor function is intrinsic to SMRT as opposed to the TR or RAR itself (Baniahmad et al., *Proc. Natl. Acad. Sci. USA* **90**:8832-8836 (1993); and Fondell et al., *Genes Dev* **7**:1400-1410 (1993)). It is demonstrated herein that the ligand triggers a dissociation of SMRT from the receptor, which would lead to an initial step in the activation process. This would be followed (or be coincident) with an induced conformational change in the carboxy-terminal transactivation domain (c, also called AF2), allowing association with co-activators on the transcription machinery (Douarin et al., *EMBO J.* **14**:2020-2033 (1995); Halachmi et al., *Science* **264**:1455-1458 (1994); Lee et al., *Nature* **374**:91-94 (1995); and Cavailles et al., *Proc. Natl. Acad. Sci. USA* **91**:10009-10013 (1994)). Thus, as has previously been suggested (Damm and Evans, (1993), *supra*), the ligand dependent activation of TR would represent two separable processes including relief of repression and net activation. The isolation of SMRT now provides a basis for dissecting the molecular basis of trans-repression.

The invention will now be described in greater detail by reference to the following non-limiting examples.

<u>Example 1</u>

<u>Isolation of SMRT</u>

Using a GAL4 DBD-RXR fusion protein (see, for example, USSN 08/177,740, incorporated by reference herein in its entirety) as a bait in a yeast

two-hybrid screening system (Durfee et al., (1993), *supra*), several cDNA clones encoding receptor interacting proteins were isolated. One of these proteins, SMRT, interacts strongly with unliganded RAR and TR but only weakly with RXR or other receptors in yeast. This protein was selected for further characterization.

Example 2

Far-western blotting procedure

Total bacteria extracts expressing GST fusions of hRARα (aa 156-462) or hRXRα LBD (aa 228-462) and control extracts expressing GST alone or GST-PML fusion protein were subjected to SDS/PAGE and electroblotted onto nitrocellulose in transfer buffer (25 mM Tris, pH 8.3/ 192 mM glycine/ 0.01% SDS). After denaturation/renaturation from 6 M to 0.187 M guanidine hydrochloride in HB buffer (25 mM HEPES, pH 7.7/25 mM NaCl/5 mM $MgCl_2$/1 mM DTT) filters were saturated at 4°C in blocking buffer (5% milk, then 1% milk in HB buffer plus 0.05% NP40). *In vitro* translated ^{35}S-labeled proteins were diluted into H buffer (20 mM Hepes, pH 7.7/75 mM KCl/0.1 mM EDTA/2.5 mM $MgCl_2$/0.05% NP40/ 1% milk/1 mM DTT) and the filters were hybridized overnight at 4°C with (1 μM) or without ligand. After three washes with H buffer, filters were dried and exposed for autoradiography or quantitated by phosphoimager.

GST-SMRT is a GST fusion of the C-SMRT encoded by the yeast two hybrid clone. GST-SMRT has been purified, but contains several degradation products.

For yeast two-hybrid screening, a construct expressing the GAL4 DBD-hRXRα LBD (aa 198-462) fusion protein was used to screen a human lymphocyte cDNA library as described (Durfee et al., (1993), *supra*). Full length SMRT cDNA was isolated from a human HeLa cDNA library (Clontech) using the two-hybrid insert as a probe.

Using the above-described far-western blotting procedure, ^{35}S-labeled SMRT preferentially complexes with bacterial extracts expressing the RAR, marginally associates with RXR and shows no association with control extracts. In contrast, ^{35}S-PPAR selectively associates with its heterodimeric partner, RXR, but not with RAR. In a similar assay, ^{35}S-labeled RAR or TR interacts strongly with SMRT and their heterodimeric partner, RXR, but not with degraded GST products, while ^{35}S-RXR interacts only weakly with SMRT. Binding of ligand to RAR or TR reduces their interactions with SMRT but not with RXR, while binding of ligand to RXR has only slight effect. Figure 1 shows the quantitation of a dose-dependent dissociation of SMRT from RAR or TR by all-*trans* retinoic acid (atRA) or thyroid hormone (triiodothyronine or T3), demonstrating that the amount of ligand required for 50% dissociation in both cases are close to the kds for both ligands (Munoz et al. *EMBO J.* **7**:155-159 (1988); Sap et al., *Nature* **340**:242-244 (1989); and Yang et al., *Proc. Natl. Acad. Sci. USA* **88**:3559-3563 (1991)).

Full length SMRT encodes a polypeptide of 1495 amino acids rich in proline and serine residues (see Figure 2 and SEQ ID NO:1). Genbank database comparison reveals similarity of the C-terminal domain of SMRT to a partial cDNA encoding another receptor interacting protein, RIP13 (Seol et al., (1995), *supra*), whose role in receptor signaling is unknown. Within this region, there can be identified several potential heptad repeats which might mediate protein-protein interaction with the "a-helical sandwich" structure (Bourguet et al., *Nature* **375**:377-382 (1995)) of the ligand binding domain (LBD) of receptors.

Example 3

Characterization of SMRT

Unlike other nuclear receptors, unliganded RAR and TR possess a strong silencing domain which represses basal level promoter activity of their target genes (Damm et al., *Nature* **339**:593-597 (1989); Brent et al., *New Biol.* **1**:329-336 (1989); Baniahmad et al., *Cell* **61**:505-514 (1990); and Baniahmad et al., *EMBO J.*

11:1015-1023 (1992)). The preferential interaction of SMRT with RAR and TR in the absence of hormone suggests that SMRT may play a role in mediating the transcriptional silencing effect of the receptor.

To further investigate the involvement of SMRT in silencing, the interaction of SMRT with mutant receptors which display distinct silencing and/or transactivation activities was tested as follows. ^{35}S-methionine labeled receptors were used as probes to hybridize immobilized GST-SMRT in the presence (10 μM) or absence of all-*trans* retinoic acid (atRA). The total bacteria extract expressing GST-RXR was included as a control.

When quantitated by phosphoimager, RAR403 shows a 4-fold better interaction with SMRT than wild type RAR. Both full length RAR or a deletion mutant expressing only the ligand binding domain (LBD, referred to as ΔΔR) associate with SMRT; this association is blocked by ligand.

These results confirm that the LBD alone is sufficient in the interaction. The carboxy-terminal deletion mutant RAR403 is a potent dominant negative repressor of basal level promoter activity of RAR target genes (Damm et al., *Proc. Natl. Acad. Sci. USA* **90**:2989-2993 (1993); Tsai and Collins, *Proc. Natl. Acad. Sci. USA* **90**:7153-7157 (1993); and Tsai et al., *Genes Dev* **6**:2258-2269 (1992)). As might be predicted from the above studies, RAR403 and its amino terminal deletion derivative, R403, interact strongly with SMRT in either the presence or absence of ligand, consistent with SMRT mediating the repressor activity of this mutant.

Example 4

Interaction of SMRT with TR Mutants

The interaction of SMRT with two different classes of TR mutants was analyzed next. The first mutant employed is the naturally occurring oncogene, v-erbA, which has strong silencing ability but no transactivation activity (Sap et al., (1989),

supra; Sap et al., *Nature* **324**:635-640 (1986); Weinberger et al., *Nature* **318**:670-672 (1985); and Weinberger et al., *Nature* **324**:641-646 (1986)). The second mutant employed is a single amino acid change (Pro 160 -> Arg) of the rTRa (TR160) which has previously been shown to lose its capacity in basal level repression but retains hormone dependent transactivation (Thompson et al., *Science* **237**:1610-1614 (1987); and Damm and Evans, *Proc. Natl. Acad. Sci. USA* **90**:10668-10672 (1993)). If SMRT is involved in silencing, it would be expected that SMRT should interact with the v-erbA, but show little or no association with the silencing-defective TR160 mutant.

Interaction of the oncogenic v-erbA and rTRα R160 mutant (TR160) with GST-SMRT was determined in a far-western assay as described above (see Example 2). When quantitated by phosphoimager, the v-erbA shows an 18-fold better interaction with SMRT than hTRβ, and the TR160 mutant shows a 10-fold lower signal than the rTRα .

As one might expect, v-erbA interacts strongly with SMRT both in presence or absence of ligand. In contrast, full length TR160 mutant or LBD of TR160 ($\Delta\Delta$TR160) does not interact significantly with SMRT when compared to the wild type receptor.

These data demonstrate that SMRT plays an important role in mediating transcriptional silencing effects of both RAR and TR. These data also suggest that the release of SMRT from receptors could be a prerequisite step in ligand-dependent transactivation by nuclear receptors.

Example 5

Formation of ternary complexes containing SMRT

RAR and TR form heterodimers with RXR, resulting in a complex with high DNA binding ability (Bugge et al., *EMBO J.* **11**:1409-1418 (1992); Yu et al., *Cell* **67**:1251-1266 (1991); and Kliewer et al., *Nature* **355**:446-449 (1992)). Since SMRT

interacts with RAR and TR, tests were conducted to determine whether SMRT can also interact with the receptor-DNA complex. Thus, the interaction of SMRT with RXR-RAR heterodimer on a DR5 element (i.e., an AGGTCA direct repeat spaced by five nucleotides) was determined in a gel retardation assay, which is carried out as follows. *In vitro* translated receptor or unprogrammed reticulocyte lysate (URL) was incubated with 1 μg of poly dIdC on ice for 15 minutes in a total volume of 20 μl containing 75 mM KCl, 7.5% glycerol, 20 mM Hepes (pH 7.5), 2 mM DTT and 0.1% NP-40, with or without ligand (in the range of about 10-100 nM employed). A ^{32}P labeled, double stranded oligonucleotide probe was added into the binding reaction (10,000 cpm per reaction), and the reaction was further incubated for 20 minutes at room temperature. The protein-DNA complex was separated on a 5% native polyacrylamide gel at 150 volts.

SMRT is seen to form a ternary complex with the RXR-RAR heterodimer on a DNA response element in the gel retardation assay. Addition of ligand releases SMRT from this complex in a dose-dependent manner.

Similarly, SMRT is seen to form a ternary complex with the RXR-TR heterodimer on a TR response element; addition of T3 disrupts the formation of this complex.

These data demonstrate that SMRT can be recruited to DNA response elements via protein-protein interaction with RAR or TR in the absence of hormone. Binding of hormone disrupts receptor-SMRT interaction and releases SMRT from the receptor-DNA complex.

Example 6

Transient transfection assay

CV-1 cells were plated in 24 well plates at a density of 50,000 cells per well. Expression plasmids were transfected into cells by lipofection using DOTAP. In

each transfection, 5 ng of GAL-RAR and 15 ng of v-erbA or SMRT were used together with 150 ng of reporter construct containing 4 copies of GAL4 binding sites in front of a minimal thymidine kinase promoter and a CMX-β-gal construct as an internal control. The relative luciferase activity was calculated by normalizing to the β-gal activity.

Example 7
Reversal of transcriptional silencing

Recently, it has been shown that over expression of RAR or TR could reverse the transcriptional silencing effect of the GAL4 DBD fusion of TR (GAL-TR) or RAR (GAL-RAR) (Baniahmad et al., *Mol Cell Biol* **15**:76-86 (1995); and Casanova et al., *Mol Cell Biol* **14**:5756-5765 (1994)), presumably by competition for a limiting amount of a SMRT co-repressor. A similar effect is observed herein when over expression of v-erbA or RAR403 mutants are shown to reverse the silencing effect of GAL-RAR and GAL-TR on the basal activity of a luciferase reporter (see Figure 3A and 3B).

In principle, over expression of SMRT should restore repressor activity when co-expressed with v-erbA or RAR403 competitors. Indeed, results presented in Figure 3C show that both the full length and the C-terminal domain of SMRT (C-SMRT) can titrate out v-erbA or RAR403 competitor activity and re-endow GAL-RAR and GAL-TR with silencing activity. In contrast, neither v-erbA nor SMRT show any effect on the transactivation activity of GAL-VP16 fusion. Thus, SMRT is able to block the titration effect of v-erbA and RAR403 and functionally replaces the putative SMRT co-repressor in this system.

Example 8
Direct recruitment of SMRT to a heterologous promoter

If SMRT is the mediator of transcription silencing of TR and RAR by interaction with template-bound unliganded receptors, then direct recruitment of SMRT

to a heterologous promoter should result in repression of basal level activity. This was tested by fusing full length SMRT to the GAL4 DBD (GAL-SMRT). The effect of the resulting fusion protein on the activity of the thymidine kinase promoter containing four GAL4 binding sites was analyzed. Figure 3D shows that GAL-SMRT, like GAL-TR, can silence basal promoter activity in a dose-dependent manner. In contrast, GAL-RXR shows no repression.

These data suggest that SMRT, when recruited to a promoter by direct DNA binding or via association with an unliganded receptor, functions as a potent transcriptional repressor.

Example 9
Cloning Of Human And Mouse SMRT co-repressors

This example describes the cloning of a full length human silencing mediator of retinoic acid and thyroid hormone receptor (SMRT co-repressor) and of two mouse SMRT isoforms, m-SMRTα and m-SMRTβ .

An examination of the previously described human SMRT co-repressor revealed that the first eight amino acids and upstream sequences were derived from a portion of ribonucleoprotein K sequence. Accordingly, a mouse spleen cDNA lambda ZAP II library (Stratagene; La Jolla CA) was screened at low stringency with a probe corresponding to approximately the 5' 1,000 base pairs (bp) of the previously identified human SMRT (s-SMRT). A 3.5 kilobase (kb) cDNA fragment was obtained that contained a unique sequence in addition to known s-SMRT sequence. The 5' end of this cDNA, and subsequently obtained clones, was used in successive rounds of screening of the mouse spleen cDNA library and a mouse brain cDNA library (Stratagene) and the full-length SMRTα isoform cDNA (SEQ ID NO: 6) and SMRTα isoform cDNA (SEQ ID NO: 10) were obtained. The mouse SMRT (m-SMRT) 5' sequence then was used at low stringency to screen a human pituitary cDNA library (Stratagene) to obtain the full-length human SMRT (h-SMRT) cDNA (SEQ ID NO: 1). All cDNA clones were

sequenced on both strands using standard methods, and have been deposited with GenBank as Accession No. AF103003 (h-SMRT; SEQ ID NOS: 3 and 5); Accession No. 113001 (m-SMRTα ; SEQ ID NOS: 6 and 7); and Accession No. 113002 (m-SMRTβ ; SEQ ID NOS: 8 and 9).

By sequentially shifting between the mouse spleen and mouse brain cDNA libraries, several clones containing a potential starting methionine and 5' untranslated region sequences were obtained. The complete polypeptide sequences of m-SMRT (SEQ ID NO: 7) and h-SMRT (SEQ ID NO: 5) are provided. In addition, a splice variant isolated from the mouse brain cDNA library encoded an m-SMRT co-repressor containing a deletion of amino acids 36 to 254 of SEQ ID NO: 7 (see SEQ ID NO: 3). The two m-SMRT co-repressors are designated SMRTα (SEQ ID NO: 7) and SMRTβ (SEQ ID NO: 9). Based on sequence similarity to N-CoR (see below), this deletion in m-SMRTβ removes the majority of the sequence in h-SMRT and m-SMRTα that is homologous to N-CoR repression domain 1 (RD1), including a portion of the Sin3A binding region.

The cloned h-SMRT (SEQ ID NO: 3) encodes a polypeptide that contains an additional 1130 amino acids at the amino terminus as compared to the previously described human SMRT co-repressor. The full length h-SMRT shares 84% identity with m-SMRTα . A comparison of h-SMRT (SEQ ID NO: 5) and N-CoR (SEQ ID NO: 11) revealed that the N-terminal extension of h-SMRT (amino acids 1 to 1030) and N-CoR (amino acids 1 to 1031) share approximately 41% identity, which is somewhat higher that the 36% identity shared between the full length proteins. However, regions within the N-CoR and SMRT N-termini share striking homology (Figures 4A and 4B).

Amino acids 1 to 160 of N-CoR are moderately conserved in h-SMRT (and m-SMRTα), sharing about 36% identity. This region of N-CoR has been reported to interact with Siah2 (Zhang et al., (1998), *supra*) and, similarly, can be involved in an

interaction of Siah2 with h-SMRT or m-SMRTα. In particular, highly conserved sequences in this region can be the specific Siah2 interaction sites (see Figure 4A).

A 52 amino acid segment from N-CoR (amino acids 255 to 312) mediates an interaction with Sin3A (Heinzel et al., *Nature* **387**:43-48 (1997)), and was presumed to represent the core of the larger RD1 region (Horlein et al., (1995), *supra*). This small interaction domain is highly conserved (about 83% identity) in h-SMRT, and the overall identity shared between SMRT and N-CoR RD1 is about 57%.

Amino acids 312 to 668 of N-CoR also are well conserved (66% identity) in h-SMRT (and m-SMRTα), and two internal blocks of sequences in this region share even greater similarity (see Figure 1B; shaded regions). These blocks are homologous to each other and to part of the SANT domain, which was identified in the yeast chromatin remodeling factor, SWI3, the yeast adapter protein, ADA2, the basal transcription factor TFIIIB, and other proteins (Aasland et al., *Trends Biochem. Sci.* **21**:87-88 (1996)), suggesting that these domains share a common and important function. The amino acids of N-CoR RD2 (see Horlein et al., (1995) *supra*) are the least conserved in h-SMRT, sharing about 30% identity.

These results demonstrate that isoforms of SMRT co-repressors are expressed in cells, as exemplified by m-SMRTα and m-SMRTβ. In addition, the results demonstrate that the previously undescribed amino terminus of SMRT co-repressors shares regions of substantial homology with N-CoR, and regions of homology are identified that indicate these sequences can mediate previously uncharacterized functions.

Example 10

Expression And Chromosomal Localization Of Smrt Co-Repressors

This example describes the tissue distribution of SMRT RNA and the chromosomal localization of human SMRT.

Total RNA was prepared from adult CB6F1 mouse tissues using TRIZOL reagent (GIBCO/BRL), and poly(A) RNA was purified from total RNA using an OLIGOTEX mRNA Kit (Qiagen, Valencia, CA). RNA was separated on 1.25% agarose/6% formaldehyde gels and transferred to a NYTRAN membrane (Scheicher & Schuell). A 720 bp m-SMRT/PstI fragment was used as a probe. Following hybridization with the SMRT probe, the filters were stripped and hybridized with a murine glyceraldehyde-3-phosphate dehydrogenase cDNA probe to allow normalization for RNA loading.

Chromosomal localization of SMRT was determined by fluorescence in situ hybridization using the 5.3 kb h-SMRT cDNA clone. The probe was labeled by nick-translation with biotin-11-dUTP, then hybridized to normal male human metaphase chromosomes. Chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI). Chromosome identification was carried out by computer inversion of the gray scale DAPI image on a PSI Imaging System (Perceptive Scientific Instruments; League City TX). Chromosome 12 confirmation was carried out using a chromosome 12-specific alpha satellite probe (Vysis; Downers Grove IL).

Previous studies using the short human SMRT co-repressor suggested that SMRT was expressed ubiquitously in various tissues. To confirm this result, expression of the full length m-SMRT was determined by northern blot analysis by using a probe consisting of nucleotides 2760 to 3620 of m-SMRT (SEQ ID NO: 6). The expression pattern was ubiquitous, as previously described, although higher levels were detected in lung, spleen, and brain. Similarly, h-SMRT was expressed ubiquitously as determined using a multiple tissue blot (CLONTECH; Palo Alto CA). It is noteworthy that two isoforms of SMRT were present in the majority of the mouse tissues and likely correspond to the m-SMRTα and m-SMRTβ isoforms.

The chromosomal location of the h-SMRT and N-CoR genes was mapped. The h-SMRT clone hybridized to the q arm of one of the C group

chromosomes. Computer-mediated banding of the DAPI stained chromosomes identified the labeled chromosome as chromosome 12, band q24. The chromosome 12 localization was confirmed by cohybridization of SMRT and a chromosome 12 alpha satellite probe, D12Z3 (Vysis), which labels the pericentromeric region of chromosome 12. The location for the human N-CoR gene was determined through a mapped human bacterial artificial chromosome clone, hCIT529I10, which is 158 kb of genomic N-CoR and resides on chromosome 11p11.2. The SMRT and N-CoR chromosomal locations can be accessed through GENEMAP98 from the Human Genome Project at http://www.ncbi.nlm.nih.gov/genemap.

These results demonstrate that the full length SMRT co-repressors and the SMRT co-repressors are expressed in various tissues. The results also demonstrate that the human SMRT gene is located on chromosome 12.

Example 11
Functional Characterization Of SMRT Amino Terminus Domains

This example demonstrates that various domains of the SMRT amino terminus can repress nuclear receptor transcriptional activity.

Experiments were performed using the plasmids pCMX-GAL4 DBD and pMH100-TK-luc (Nagy et al., (1997), *supra*). Standard PCR amplifications were used to generate GAL4 fusion constructs. All constructs were verified by double-stranded sequencing to confirm identity and reading frame.

Monkey CV-1 cells were grown in DMEM supplemented with 10% resin-charcoal stripped fetal bovine serum (FBS), 50 units/ml of penicillin G, and 50 μg/ml of streptomycin sulfate at 37°C in 7% CO_2. V-1 cells (60-70% confluence, 48-well plate) were cotransfected with 16 ng of pCMX-GAL4, 100 ng of pMH100-TK-luc, and 100 ng of pCMX-β galactosidase in 200 μl of DMEM containing 10% super-

stripped fetal calf serum (FCS) by the N-(1-(2,3-dioleoyloxy)propyl)-N,N,N-trimethylammonium methylsulfate (DOTAP)-mediated procedure (Nagy et al., (1997), *supra*). The amount of DNA in each transfection was kept constant by addition of pCMX. After 24 hr, the medium was replaced; cells were harvested and assayed for luciferase activity 36 to 48 hr after transfection. Luciferase activity was normalized by the level of β-galactosidase activity. Each transfection was performed in triplicate and repeated at least three times.

Based on the high degree of identity between regions of the SMRT amino terminus and the corresponding N-CoR region, the ability of regions in the SMRT amino terminus to act in transcriptional repression was examined. A nested series of nucleotide sequences encoding portions of the SMRT amino terminus fused to the GAL4 DNA binding domain (GAL-DBD) was prepared in mammalian expression vectors (Figure 5A). The constructs were cotransfected with a GAL4-TK-luciferase reporter plasmid to determine the regulatory properties of the GAL4-SMRT fusions. Repression was determined relative to the basal activity of the reporter in the presence of the GAL-DBD alone.

The entire SMRT amino terminus region (GAL4-SMRT(1-1031)) demonstrated the greatest amount of repression (approximately 38-fold), and virtually extinguished reporter activity. In comparison, GAL4-SMRT (1-303), which is equivalent to N-CoR RD1, demonstrated 6-fold repression; and GAL4-SMRT (736-1031), which is equivalent to N-CoR RD2, demonstrated about 2.6-fold repression. Surprisingly, the highly conserved SANT domain conferred a significant amount of repression (about 3.3-fold).

A smaller region (amino acids 845 to 986) within the RD2 homology region shows a higher level of sequence conservation as compared to the entire RD2 region. Deletion constructs were generated to determine whether this minimal region was sufficient for the repression activity of RD2. Deletion of flanking amino acids 736 to 845 or of amino acids 987 to 1055 did not affect the level of repression, demonstrating

that the repressor function of RD2 is contained within a 141 amino acid core sequence of RD2.

Based on sequence similarity to N-CoR, the deletion of amino acids 36 to 254 in the m-SMRTβ isoform removes the majority of RD1, including a portion of the Sin3A binding region. The effect of this deletion on SMRT function was examined by cotransfection experiments comparing repression by SMRTα to SMRTβ. These experiments revealed that SMRTβ has substantially less repressor activity than SMRTα. A construct containing the entire amino terminus of m-SMRTβ (amino acids 1 to 813) repressed activity about 2.6 fold, as compared to m-SMRTα amino acids 1 to 1031, which repressed activity about 38.1-fold. In addition, a GAL4 construct containing m-SMRT amino acids 1 to 83 repressed activity only about 1.4-fold. These results indicate that alternative splicing can add further diversity to expand the function of SMRT gene products.

Example 12

Yeast Two-Hybrid Screen and Assays

To investigate whether repression by EcR in CV-1 cells is mediated by its association with a vertebrate corepressor and whether such an interaction, if it does occur, is impaired by the A483T mutation, a mammalian two-hybrid assay with Gal4-c-SMRT was conducted.

A yeast two-hybrid screen (Fields and Song, *Nature*, **340**:245-246, (1989)) was performed by transforming approximately 2 X 10^6 Y190 yeast cells with a pAS-EcR construct and a Drosophila (0-8 hr) embryonic c-DNA two-hybrid library (Yu et al., *Nature*, **385**:552-555, (1997)). Transformants were selected onto DO-Leu-Trp-His plates containing 40 mM 3-aminotriazole (Sigma) for 3-4 days. Surviving yeast colonies were picked as primary positives and restreaked on selection plates to isolate single clones. Activation domain plasmids were rescued from the selected positive transformants for further analysis. Each clone was evaluated by testing its

potential interaction with several other nuclear receptors using the yeast two-hybrid assays. E52 was isolated and further pursued based on this selection criterion. Quantitative liquid assay of β-galactosidase was performed on positive clones 16 hr after treating the yeast cells with no ligand, or with 3 μM ligand.

pAS-EcR is a fusion gene with the region corresponding to amino acids 223-878 of EcRB1 fused C-terminally to the Gal4-DBD of the pAS1-CYH2 construct (Durfee et al., (1993), *supra*); other Gal4-DBD-based nuclear receptor constructs used in this yeast two-hybrid assay include: USP (amino acids 50-508), hRAR (amino acids 186-462) and hTR (amino acids 121-410) (Schulman et al., (1995), *supra*), and SMRT (Chen and Evans, (1995), *supra*). β-galactosidase activities were quantified by liquid assay for yeast cells treated either without ligand or with 3 μM of corresponding hormone. All-trans retinoic acid (ATRA) is a ligand of RAR; 3,3',5-triiodothyroacetic acid (T3) is a ligand of TR.

Similar yeast two-hybrid assays were also used to examine the interaction between SMRTER and mSin3A and dSin3A.

Example 12
Cloning SMRTER

To isolate full-length SMRTER cDNA, a XhoI insert fragment isolated from the E52 clone was used to screen male and female Tudor c-DNA libraries (gift of Tulle Hazelrigg). This initial screen resulted in isolating three overlapping c-DNA clones covering the region of amino acid 2094 to the C terminus of SMRTER Additional regions were obtained from three consecutive library screens using two cosmid clones isolated from the Tamkun genomic library (gift of John Tamkun). Sequences of these overlapping c-DNA and genomic clones were assembled to obtain a conceptual open reading frame of SMRTER 3446 amino acids in length (SEQ ID NO:12; Figure 8A). The translational initiation codon was designated based on the sequences that match the consensus Kozak codons and is preceded by three in-frame

consecutive stop codons in the upstream region. Both strands of the sequences of the c-DNA clones were determined using an ABI prism Big Dye® terminator cycle sequencing ready reaction kit (PE Biosystems) and ABI 377 instrument.

Example 14
Plasmids

CMV promoter-driven expression plasmids of EcR, USP, RXR, c-SMRT, β-galactosidase, and pMH100-TK-luc reporter, and yeast plasmids of RAR, TR, and SMRT have been described previously (Yao et al., (1992), *supra*, Yao et al., (1993), *supra*; Chen and Evans, (1995), *supra*; Schulman et al., (1995), *supra*; Chen et al., *Proc. Natl. Acad. Sci. USA* **93**:7567-7571, (1996); Nagy et al., (1997), *supra*). hsp27EcR-TK-Luc, a reporter with six copies of the hsp27EcRE, is a gift of Barry Forman. CMV vector-driven EcR A483T and Gal4-SMRD3 mutations were generated using the Transformer® site-directed mutagenesis kit (Clontech) with proper selection primers and the mutagenic primers that correspond to the missense mutation (A483T) of EcR and to the designated mutations, M2 and M3, in the SMRD3 domain, respectively. Other plasmids were constructed with standard techniques, including various enzyme digestions or PCR amplification.

Example 15
Cell Culture and Transfection

CV-1 cells were grown in Dulbecco's modified Eagles medium at 37°C in 5% CO_2. The media were supplemented with 10% AG1-X8 resin charcoal double-stripped calf bovine serum, 50 U/ml penicillin G, and 50 μg/ml streptomycin sulfate. Approximately 20 hr after CV-1 cells (10^5 cells) were plated in 48-well cell culture clusters (Costar), cells were transiently transfected with plasmids using DOTAP according to the instructions of the manufacturer (Boehringer Mannheim). The amount of CMV promoter-driven expression vectors, β-galactosidase gene

expression vector, CMX-lacZ, and reporter, pMH100-TK-luc or hsp27EcRE-TK-Luc, were in the range of 100-200 ng, 500 ng, and 400 ng, respectively, for six wells of each 48-well clusters in each transfection experiments. At least 4 hr after transfection, each medium was replaced with medium either without ligand, or with 1 μM of MurA. Cells were harvested and assayed approximately 48 hr after transfection. All experiments were performed in triplicate and repeated with similar results.

CV-1 cells were transfected with wild-type EcR or EcR A483T, along with vp16-USP and a reporter, hsp27EcRE-TK-Luc, which contains six copies of the hsp27EcRE fused to the thymidine kinase (TK) promoter-luciferase reporter. VP16-USP fusion contains the region of USP (amino acids 50-508) fused C-terminally to the VP16 domain. Muristerone A (MurA) is a potent ecdysone agonist (Christopherson et al., *Proc. Natl. Acad. Sci. USA,* **89**:6314-6318, (1992)). In all experiments, cells were also cotransfected with CMV-lacZ, which is used to normalize the luciferase activity. As shown in Figure 6A, the ability to dimerize with USP is reflected in reporter activity without treatment with hormone (open bar), and the ability to respond to hormone is reflected in reporter activity when cells were treated with 1 μM Muristerone A (closed bar).

CMV promoter-driven expression vector including wild-type EcR or EcR A483T was cotransfected with VP16-USP and Gal4-c-SMRT (amino acids 981 to C terminus) (Chen and Evans, (1995), *supra*) into CV-1 cells to examine its effect on the interaction with vertebrate corepressor. All cells were also cotransfected with a TK-luciferase reporter construct, pMH100-TK-Luc, containing four copies of the yeast Gal4-responsive element. EcR A483T corresponds to a single amino acid change (alanine→threonine) at the 483 site of EcR (Bender et al., (1997), *supra*). The results of this experiment (Figure 6B) show that EcR A483T disrupts the interaction with SMRT.

Example 16

In Vitro Interacting Assays

Glutathione S-transferase fusion proteins, including GST-X, GST-ERID1 (amino acids 1698-2063 of SEQ ID NO:1), and GST-ERID2 (amino acids 2951-3038 of SEQ ID NO:1), were expressed in E. coli DH5 cells, and extracts were affinity purified by binding to glutathione Sepharose 4B beads. Bound proteins used as affinity matrices in pull-down experiments were first equilibrated with the binding buffer (20 mM HEPES [pH 7.9], 150 mM NaCl, 1 mM EDTA, 4 mM MgCl2, 1 mM DTT, 0.06% NP40, 10% Glycerol, 0.25 mM PMSF, 1 mg BSA). For pull-down assays using GST-ERID1 (amino acids 1698-2063 of SEQ ID NO:1) and GST-ERID2 (amino acids 2951-3038 of SEQ ID NO:1), additional hsp27EcRE (0.05 μg/ml) was added to the binding buffer. In this experiment, 30 μl of 50% GST-protein beads slurry, containing approximately 1 μg of proteins, were incubated with 10 μl of 35S-methionine-labeled proteins in 300 μl of the binding buffer (with or without 3 μM of MurA as indicated) for 30 min at room temperature. After the incubation, beads were washed three times with the binding buffer (with or without ligand) and resuspended in SDS-PAGE sample buffer before loading. After electrophoresis, bound radio-labeled proteins were visualized by autoradiography. 35S-methionine-labeled EcR, USP were generated in a coupled transcription-translation system, TNT (Promega), using CMX-EcR (T7) and CMX-uspK (T7) constructs as templates, respectively.

Example 17

Immunohistochemistry and Immunofluorescence

Antibodies against SMRTER were raised in rabbits immunized with bacterially expressed glutathione S-transferase fusion proteins corresponding to the region (amino acids 2477-2648 of SEQ ID NO:1) of SMRTER. Specific antibodies were purified by affinity chromatography through antigen-linked columns and used at 1:200 dilution for tissue staining. Tissues for whole-mount staining were dissected at the wandering third instar stage of the Canton-S strain larvae and fixed (4%

formaldehyde in 1? PBS, 50 mM EGTA) for at least 30 min. Preincubation, secondary antibodies, washes, and peroxidase reactions are described in the protocol of the Elite ABC (Rabbit IgG) kit (Vector). For the pilot experiments, partially purified IgG from preimmunization serum was used. For polytene chromosome staining, salivary glands were dissected according to the method described in Zink et al., *EMBO J.*, **10**:153-162, (1991).

Chromosome spreads were costained with affinity-purified anti-SMRTER (1:100) polyclonal antibody and with anti-USP monoclonal antibody (ABIII/AD5; gift of F. Kafatos, 1:100 dilution). SMRTER was detected with Texas red-conjugated donkey anti-rabbit secondary antibody (1:100 dilution), and USP was detected with FITC-conjugated donkey anti-mouse secondary antibody (1:100 dilution) (Jackson ImmunoResearch Labs).

Example 18
ER Interacts Genetically with DSinA

In keeping with the evidence that dSin3A is a component in EcR regulatory pathway, an experiment was conducted to examine whether dSin3A interacts genetically with EcR using several previously characterized Drosophila EcR and dSin3A mutants (Bender et al., (1997), *supra*; Neufeld et al., (1998), *supra*). In the experiment, in which female dSin3AK07401 were crossed with male EcRE261st using techniques known in the art (see Table 1 below), only approximately 14% of the scored EcRE261st/dSin3AK07401 progenies survived, a percent that is significantly lower than the expected 33.3%. This suggests that a large portion of the EcRE261st/dSin3AK07401 flies either die prior to eclosion or fail to eclose. Additionally, surviving EcRE261st/dSin3AK07401 escapers showed delayed development and wing defects, in which wings are held horizontally at 45°-90° angle from the body axis. These results suggest that dSin3A shares an overlapping regulatory pathway with EcR.

In a reverse genetic cross, in which female EcRE261st were crossed with male dSin3AK07401, none of the EcRE261st/dSin3AK07401 flies survived to adulthood. These results suggest that EcRE261st/dSin3AK07401 results in a genetically sensitized background. When the maternally deposited EcR in embryos descended from female EcRE261st/SM6b was cut in half, the lethality for EcRE261st/dSin3AK07401 was further increased. These results reveal that, in addition to its previously known zygotic function, EcR also contributes maternally to Drosophila development.

Table 1

Table 1. EcR Interacts Genetically with DSin3A

Cross		$EcR^{E261st}/DSin3A^{KO7401}$ Surviving Rate (%)
$DSin3A^{KO7401}/CyO$ ♀ × $EcR^{261st}/SM6b$ ♂		14 (n = 141)
$EcR^{261st}/SM6b$ ♀ × $DSin3A^{KO7401}/CyO$ ♂		0 (n = 144)

A similar wing held-out phenotype is also observed in $EcR^{E261st}/DSin3A^{xe374}$, Df(2R)nap11/$DSin3A^{KO7401}$, and Df(2R)nap11/$Dsin2A^{xe374}$. EcR^{E261st} and Df(2R)nap11 are both described in Figure 6. $Dsin2A^{KO7401}$ is an allele with a P element insertion within the 5' intron of Sin3A. $DSin3A^{xe374}$ is an X ray-generated allele (Neufeld et al., (1998)). n=the number of surviving flies scored. Note that CyO/SM6b is lethal.

EcRA483T flies showed developmental abnormalities in wings and tergites. A similar phenotype, although with a lower penetration rate, has been also observed in EcRA483T/Df(2R)20B and in EcRA483T/Df(2R)nap11. Df(2)20B and Df(2)nap11 are both deficiencies in which EcR is deleted (Bender et al., (1997), *supra*). Sequence alignment of EcR with the vertebrate TR, RAR, and v-erbA, an oncogenic TR variant, revealed that alanine 483 is located within a highly conserved 23-amino acid (aa) loop region connecting helices 3 and 4, termed the LBD signature

motif (Wurtz et al., *Nat. Struct. Biol.*, **3**:206, (1996)) (see Figure 6C). Based on structural studies of vertebrate nuclear receptors (for review, see Moras and Gronemeyer, (1998), *supra*), this alanine residue appears to be on the exposed surface, consistent with it being a potential corepressor binding site for nuclear receptors.

These in vivo studies indicate that EcRA483T is a semilethal allele (Bender et al., (1997), *supra*). When EcRA483T is in trans with EcRE261st, an allele that removes both the DBD and LBD domains of EcR, animals are primarily lethal (>95%). The few surviving EcRA483T/EcRE261st flies, however, display significant delays in development, blistered wings, and defective tergites, indicating that EcR is involved in the development of these tissues. The ability of EcR to bind a vertebrate corepressor and the loss of this property in EcR A483T suggests that the defects observed in EcRA483T flies may result from the disruption of its interaction with an as yet unidentified Drosophila corepressor.

Example 19

Isolation of an EcR-Interacting Factor

The CMV promoter-driven expression vector including wild-type EcR or EcR A483T, was cotransfected with vp16-USP and Gal4-c-SMRT (amino acids 981 to C terminus) (Chen and Evans, (1995), *supra*) into CV-1 cells to examine its effect on the interaction of the invertebrate SMRTER with vertebrate corepressor. All cells were also cotransfected with a TK-luciferase reporter construct, pMH100-TK-Luc, containing four copies of the yeast Gal4-responsive element. EcR A483T corresponds to a single amino acid change (alanine→threonine) at the 483 site of EcR (Bender et al., (1997), *supra*). Although EcR readily interacted with SMRT in both mammalian and yeast cells (Figure 6B; Figure 7), repeated low-stringency hybridization screens failed to identify a Drosophila homolog of SMRT. No SMRT/N-CoR homolog was found in C. elegans.

Example 20

Isolation and Characterization of an EcR-Interacting Clone - Yeast Two-hybrid screen

To pursue the isolation of an EcR corepressor, a yeast two hybrid interaction screen was performed of a Drosophila embryonic cDNA library using pAS-EcR as bait. E52 was isolated as one of the complementary positive clones from a yeast two-hybrid screen with pAS-EcR as bait, as described in Example 12.

Example 21

Characterization of a Repression-Defective EcR Allele, EcRA483T

(A) CV-1 cells were transfected with wild-type EcR or EcR A483T, along with vp16-USP and a reporter, hsp27EcRE-TK-Luc, which contains six copies of the hsp27EcRE fused to the thymidine kinase (TK) promoter-luciferase reporter. In all experiments, cells were also cotransfected with CMV-lacZ, which is used to normalize the luciferase activity. The ability to dimerize with USP was reflected in reporter activity without treatment with hormone (open bar), and the ability to respond to hormone was reflected in reporter activity when cells were treated with 1 μM Muristerone A (closed bar). vp16-USP fusion contains the region of USP (amino acids 50-508) fused C-terminally to the vp16 domain. Muristerone A (MurA) is a potent ecdysone agonist (Christopherson et al., (1992), *supra*). In these tests EcR A483T was selectively defective in repression.

(B) CMV promoter-driven expression vector including wild-type EcR or EcR A483T was cotransfected with vp16-USP and Gal4-c-SMRT (amino acids 981 to C terminus) (Chen and Evans, (1995), *supra*) into CV-1 cells to examine its effect on the interaction with vertebrate corepressor. All cells were also cotransfected with a TK-luciferase reporter construct, pMH100-TK-Luc, containing four copies of the yeast Gal4-responsive element. EcR A483T corresponds to a single amino acid change (alanine threonine) at the 483 site of EcR (Bender et al., (1997), *supra*). The

results of this test show that EcR A483T disrupts the interaction with SMRT.

(C) Sequence alignment of EcR with the vertebrate TR, RAR, and v-erbA, an oncogenic TR variant, reveals that the alanine 483 of the EcRA4831T mutant is located within a highly conserved 23-amino acid (aa) loop region connecting helices 3 and 4, termed the LBD signature motif (Wurtz et al., (1996), *supra*) (Figure 6C). Based on structural studies of vertebrate nuclear receptors (for review, see Moras and Gronemeyer, (1998), *supra*), this alanine residue appears to be on the exposed surface, consistent with it being a potential corepressor binding site for nuclear receptors.

In vivo studies indicated that EcRA483T is a semilethal allele (Bender et al., (1997), *supra*). When EcRA483T is in trans with EcRE261st, an allele that removes both the DBD and LBD domains of EcR, animals are primarily lethal (>95%). The few surviving EcRA483T/EcRE261st flies, however, display significant delays in development, blistered wings, and defective tergites, indicating that EcR is involved in the development of these tissues. The ability of EcR to bind a vertebrate corepressor and the loss of this property in EcR A483T suggested to us that the defects observed in EcRA483T flies may result from the disruption of its interaction with an as yet unidentified Drosophila corepressor.

Example 22

Isolation of an EcR-Interacting Factor

Although EcR readily interacts with SMRT in both mammalian and yeast cells (Figure 6B; Figure 7), repeated low-stringency hybridization screens failed to identify a Drosophila homolog of SMRT. Given that no SMRT/N-CoR homolog is found in C. elegans, it was believed that either a SMRT/N-CoR-like corepressor is not conserved in invertebrates or, alternatively, invertebrate corepressors may diverge significantly from their vertebrate counterparts. To pursue the isolation of an EcR corepressor, a yeast interaction screen of a Drosophila embryonic cDNA library using

EcR as bait was conducted as described in Example 19. This screen resulted in the isolation of a clone, E52, whose protein product interacts with EcR as well as with the vertebrate RAR and TR, but notably not with USP (Figure 7). Unlike the interaction between E52 and RAR, which can be dissociated by all-trans retinoic acid, the interaction between E52 and EcR, or the interaction between SMRT and EcR, is not dissociated by Muristerone A (MurA). This result suggests that other factors essential for the dissociation of E52 from EcR, such as USP, are missing in yeast (see below).

Example 23

Isolation and Characterization of an EcR-Interacting Clone

E52 was isolated as one of the complementary positive clones from a yeast two-hybrid screen. Isolation of overlapping cDNA and genomic clones led to the identification of a full-length sequence encoding a large protein of 3446 amino acids (Figure 8A). This protein contains several unusually long stretches of Gln, Ala, Gly, and Ser repeats. Comparative analysis reveals it to be a novel protein with limited regions of clear homology with the vertebrate nuclear receptor corepressors SMRT and N-CoR (Chen and Evans, (1995), *supra*; Hörlein et al., (1995), *supra*; Ordentlich et al., (1999), *supra*; Park et al., (1999), *supra*). This protein SMRTER, SMRT-related ecdysone receptor-interacting factor, was shown by Northern blot analysis to encode large transcripts (>12 kb) expressed broadly throughout the embryonic stage and three larvae stages, as well as in adult Drosophila flies.

Example 24

Molecular and Biochemical Analysis for ERID1 and ERID2

Interaction with the EcR complex was evaluated based on transient transfection with the Gal4-SMRTER fusion genes. USP, EcR-vp16 (VP16 transactivating domain was fused C-terminally to the end of the EcRB1 isoform), and the reporter, pMH100-TK-Luc.

In vitro pull down assays (Example 12) were conducted to determine whether EcR interacts with ERID1 and ERID2. In vitro translated 35S-methionine-labeled EcRB1 alone, or a mixture of 35S-methionine-labeled EcRB1 and unlabeled USP, or 35S-methionine-labeled USP alone, were incubated with GST, GST-ERID1 (amino acids 1698-2063 of SEQ ID NO:1), or GST-ERID2 (amino acids of SEQ ID NO:1). GST-ERID1 and GST-ERID2, but not GST alone, pull down labeled EcR, whereas little interaction is found between USP and any of the three GST proteins. In addition, the pull-down complex was disrupted by the addition of 3μM MurA when USP is present. These in vitro results establish that SMRTER and EcR may interact directly.

Further in vitro tests were conducted to determine ERID1, ERID2, and c-SMRT compete with each other to bind EcR. Gal4-ERID1 (amino acids 1698-2063 of SEQ ID NO:1) or Gal4-ERID2 (amino acids 2929-3181 of SEQ ID NO:1), along with EcR-vp16 and USP, were transfected in CV-1 cells as described above. In this competition experiment, additional ERID1, ERID2, and c-SMRT (Chen et al., (1996), *supra*) were cotransfected into cells. ERID1 (1698-2063) and ERID2 ((amino acids 2929-3038 of SEQ ID NO:1) were tagged with the nuclear targeting signal (MAPKKKRKV) (SEQ ID NO:3) to ensure that these proteins were localized in nuclei. As shown in Figure 11C, interaction between each Gal4-ERID fusion and EcR-vp16:USP was significantly decreased by both ERIDs and by c-SMRT. Interestingly, a more prominent effect was observed in experiments when Gal4-ERID1 (amino acids 1698-2063 of SEQ ID NO:1) was challenged by ERID2, and, conversely, a more efficient competition was achieved by ERID1 to Gal4-ERID2 (amino acids 2094-3181 of SEQ ID NO:1). Together, these results suggest that ERID1, ERID2, and c-SMRT may bind similar or overlapping surface(s) in EcR.

Example 25

SMRTER Colocalizes with the EcR on Polytene Chromosomes

SMRTER antibodies were prepared as described in Example 12 to examine its cytological and chromosomal localization patterns of SMRTER. Consistent with its action as a corepressor of EcR, SMRTER was localized to nuclei of salivary glands and of fat bodies, as well as to nuclei of eye, wing, and leg imaginal discs isolated from the third instar larvae.

Next association of SMRTER with the EcR:USP complex on chromosomes was examined. The USP staining pattern was used as an index for EcRs presence on chromosomes. Since USP and EcR colocalized with each other on polytene chromosomes (Yao et al., (1993), *supra*), chromosomal spreads prepared from the salivary glands of wandering third instar larvae (prior to pupariation) were subjected to simultaneous immunological staining with antibodies against SMRTER and USP. SMRTER was detected with antibody conjugated with Texas red, USP with FITC.

To visualize the band, interband, and puffing patterns of the polytene chromosomes, the chromosomes were counterstained with DAPI to show the banding regions while leaving the interbands and puffs unstained or lightly stained. Indirect immunofluorescence staining revealed that SMRTER is a chromosome-bound protein and colocalizes with USP (FITC) at a majority of chromosomal sites; whereas in a pilot experiment, no such staining patterns were detected using the preimmunization serum. The strongest SMRTER staining was primarily associated with the boundary between band and interband regions as well as within the interband regions of chromosomes counterstained with DAPI. This result confirms that, as an EcR-associating factor, SMRTER is recruited by the EcR:USP heterodimers to their specific target chromosomal loci.

SMRTER staining can still be detected in puffed regions, such as the 2B puff. Since the polytene chromosomes consist of a parallel arrangement of several hundred to two thousand copies of the euchromatic portions of the chromosomes, an individual binding protein like SMRTER may be cycling on and off, resulting in a steady state of signals detected in the broader chromatin regions. Whether or not SMRTER levels actually change prior to or after the peak of ecdysone pulses remains to be established.

While the invention has been described in detail with reference to certain preferred embodiments thereof, it will be understood that modifications and variations are within the spirit and scope of that which is described and claimed.

That which is claimed is:

1. An isolated polynucleotide encoding a member of a family of silencing mediators of retinoic acid receptor and thyroid hormone receptor, or an isoform or peptide portion thereof (SMRT co-repressor), or an isolated polynucleotide complementary thereto.

2. The polynucleotide of claim 1, which modulates transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor).

3. The polynucleotide of claim 2, wherein the SMRT co-repressor comprises a repression domain having

a) less than about 83% identity with a Sin3A interaction domain of N-CoR set forth as amino acids 255 to 312 of SEQ ID NO: 11;

b) less than about 57% identity with repression domain 1 of N-CoR set forth as amino acids 1 to 312 of SEQ ID NO: 11;

c) less than about 66% identity with a SANT domain of N-CoR set forth as amino acids 312 to 668 of SEQ ID NO: 11; or

d) less than about 30% identity with repression domain 2 of N-CoR set forth as amino acids 736 to 1031 of SEQ ID NO: 11,

and polynucleotides that hybridize thereto under stringent conditions.

4. The polynucleotide of claim 1, wherein the SMRT co-repressor is a human SMRT co-repressor having an amino acid sequence as set forth in SEQ ID NO: 5 or conservative variations thereof.

5. A polynucleotide which hybridizes under stringent conditions with a polynucleotide according to claim 2.

6. A polynucleotide that has at least 80% sequence identity with a polynucleotide according to claim 2.

7. The polynucleotide of claim 4, which has a nucleotide sequence as set forth in SEQ ID NO: 4, and conservative variations thereof.

8. The polynucleotide of claim 1, wherein the SMRT co-repressor is a mouse SMRTα isoform.

9. The polynucleotide of claim 6, having an amino acid sequence as set forth in SEQ ID NO: 7 or conservative variations thereof.

10. The polynucleotide of claim 4, which has a nucleotide sequence as set forth in SEQ ID NO: 6.

11. The polynucleotide of claim 1, wherein the SMRT co-repressor is a mouse SMRTβ isoform.

12. The polynucleotide of claim 11, having an amino acid sequence as set forth in SEQ ID NO: 9 or conservative variations thereof.

13. The polynucleotide of claim 11, which has a nucleotide sequence as set forth in SEQ ID NO: 8.

14. The polynucleotide of claim 1, comprising a nucleotide sequence selected from the group consisting of:

nucleotides 1 to 3094 of SEQ ID NO: 4;

nucleotides 1 to 3718 of SEQ ID NO: 6; and

nucleotides 1 to 2801 of SEQ ID NO: 8.

15. A polynucleotide that under stringent conditions with a polynucleotide according to claim 14, provided that the polynucleotide does not contain a sequence identical to SEQ ID NO: 11.

16. A polynucleotide that has at least 80% sequence identity with a polynucleotide according to claim 14, provided that the polynucleotide does not contain a sequence identical to SEQ ID NO: 11.

17. A polynucleotide of claim 1, comprising a nucleotide sequence selected from the group consisting of:

nucleotides 1 to 8388 of SEQ ID NO: 6; and

nucleotides 1 to 7465 of SEQ ID NO: 8.

18. The polynucleotide of claim 1, comprising nucleotides 1 to 8561 of SEQ ID NO: 4.

19. The polynucleotide of claim 1, which is operably linked to a second nucleotide sequence.

20. The polynucleotide of claim 19, which encodes a fusion polypeptide comprising the SMRT co-repressor operably linked to a DNA binding domain of a transcription factor.

21. A vector comprising the polynucleotide of claim 1.

22. A host cell containing the polynucleotide of claim 1.

23. An isolated oligonucleotide, comprising at least 15 nucleotides that can hybridize specifically to the polynucleotide of claim 1, but not to a polynucleotide encoding SEQ ID NO: 11 or to a polynucleotide encoding an amino acid sequence consisting of amino acids 1031 to 2517 of SEQ ID NO: 5.

24. The oligonucleotide of claim 23, wherein the polynucleotide encodes at least five contiguous amino acids of a sequence selected from the group consisting of:

amino acids 720 to 745 of SEQ ID NO: 5;
amino acids 716 to 742 of SEQ ID NO: 7; and
amino acids 497 to 523 of SEQ ID NO: 9.

25. The oligonucleotide of claim 23, which can hybridize specifically to a polynucleotide encoding SEQ ID NO: 5 or SEQ ID NO: 7, but not to a polynucleotide encoding SEQ ID NO: 9.

26. An isolated silencing mediator of retinoic acid and thyroid hormone receptor, or isoform or peptide portion thereof (SMRT co-repressor), wherein the co-repressor modulates transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor).

27. An isolated co-repressor comprising a repression domain having

a) less than about 83% identity with a Sin3A interaction domain of N-CoR set forth as amino acids 255 to 312 of SEQ ID NO: 11;

b) less than about 57% identity with repression domain 1 of N-CoR set forth as amino acids 1 to 312 of SEQ ID NO: 11;

c) less than about 66% identity with a SANT domain of N-CoR set forth as amino acids 312 to 668 of SEQ ID NO: 11; or

d) less than about 30% identity with repression domain 2 of N-CoR set forth as amino acids 736 to 1031 of SEQ ID NO: 11.

28. An isolated peptide, comprising at least six contiguous amino acids of an amino acid sequence selected from the group consisting of:

amino acids 1 to 1030 of SEQ ID NO: 5;

amino acids 1 to 1029 of SEQ ID NO: 7;

amino acids 1 to 809 of SEQ ID NO: 9;

and conservative variations thereof,

provided the peptide is not identical to a sequence of SEQ ID NO: 11.

29. An isolated antibody that binds specifically to the peptide of claim 28.

30. A cell line, which produces the antibody of claim 29.

31. A chimeric molecule, comprising the SMRT co-repressor of claim 26 and at least a second molecule.

32. A complex, comprising a SMRT co-repressor of claim 26 and a member of the nuclear receptor superfamily (nuclear receptor).

33. The complex of claim 32, wherein the nuclear receptor is in the form of a dimer.

34. A method for identifying an agent that modulates the repressor potential of a SMRT co-repressor, the method comprising:

a) contacting a host cell with an agent,

wherein the host cell contains a first expressible nucleotide sequence operably linked to a first DNA regulatory element, and

expresses a fusion polypeptide comprising a SMRT co-repressor of claim 26, and a DNA binding domain of a first transcription factor, which can specifically bind the first DNA regulatory element,

and wherein binding of the DNA binding domain of the first transcription factor to the first DNA regulatory element results in expression of the first expressible nucleotide sequence; and

b) detecting a change in the level of expression of the first expressible nucleotide sequence due to contacting the host cell with the agent, thereby identifying an agent that modulates the repressor potential of a SMRT co-repressor.

35. A method for identifying an agent that modulates a function of a SMRT co-repressor, the method comprising:

a) contacting a SMRT co-repressor of claim 26,

a member of the nuclear receptor superfamily (nuclear receptor), and

an agent; and

b) detecting an altered activity of the SMRT co-repressor in the presence of the agent as compared to the absence of the agent, thereby identifying an agent that modulates a function of the SMRT co-repressor.

36. A method of modulating the transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor) in a cell, the method comprising introducing a polynucleotide of claim 1 into the cell, whereby the polynucleotide or an expression product of the polynucleotide alters the level of a SMRT co-repressor in the cell, thereby modulating the transcriptional potential of the nuclear receptor.

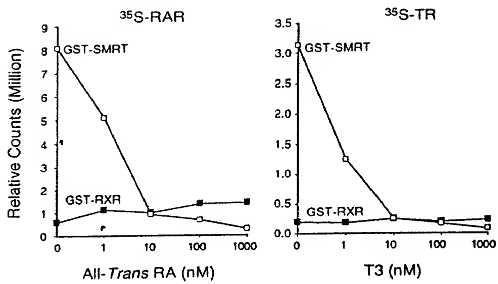
37. A method of identifying a molecule that interacts specifically with a SMRT co-repressor, the method comprising:

a) contacting the molecule with the SMRT co-repressor of claim 26; and

b) detecting specific binding of the molecule to the SMRT co-repressor, thereby identifying a molecule that interacts specifically with a SMRT co-repressor.

ABSTRACT OF THE INVENTION

The present invention relates to isolated polynucleotides encoding a family of silencing mediators of retinoic acid and thyroid hormone receptor (SMRT) isoforms, including vertebrate and invertebrate isoforms thereof. For example, a full length human SMRT co-repressor, two isoforms of a mouse SMRT-- a longer form, mouse SMRTα, and a shorter form, mouse SMRTβ, and an isoform of an insect (Drosophilia), SMRTER -- as well as peptide portions of the SMRT co-repressors that can modulate transcriptional potential of a member of the nuclear receptor superfamily (nuclear receptor); to oligonucleotides that can hybridize specifically to such a polynucleotide; to vectors and to host cells containing such polynucleotides. The invention also relates to polypeptide SMRT co-repressors encoded by such invention SMRT polynucleotides, and to peptide portions thereof that can modulate transcriptional potential of a nuclear receptor; including peptide portions of a SMRT co-repressor that are not present in an N-CoR polypeptide. In addition, the invention relates to chimeric molecules and to complexes containing a SMRT co-repressor or peptide portion thereof, to antibodies that specifically bind such compositions, and to methods for identifying an agent that modulates the repressor potential of a SMRT co-repressor. The invention also provides methods for identifying an agent that modulates a function of a SMRT co-repressor; for modulating the transcriptional potential of a nuclear receptor in a cell using the compositions of the invention; and for identifying a molecule that interacts specifically with a SMRT co-repressor.



35S-RAR
35S-TR
Relative Counts (Million)
GST-SMRT
GST-RXR
All-Trans RA (nM)
T3 (nM)

FIGURE 1

1 MEAWDAHPDKEAFAAEAQKLPGDPPCWTSGLPFPVPPREVIKASPHAPDP

51 SAFSYAPPGHPLPLGLHDTARPVLPRPPTISNPPPLISSAKHPSVLERQI

101 GAISQGMSVQLHVPYSEHAKAPVGPVTMGLPLPMDPKKLAPFSGVKQEQL

151 SPRGQAGPPESLGVPTAQEASVLRGTALGSVPGGSITKGIPSTRVPSDSA

201 ITYRGSITHGTPADVLYKGTITRIIGEDSPSRLDRGREDSLPKGHVIYEG

251 KKGHVLSYEGGMSVTQCSKEDGRSSSGPPHETAAPKRTYDMMEGRVGRAI

301 SSASIEGLMGRAIPPERHSPHHLKEQHHIRGSITQGIPRSYVEAQEDYLR

351 REAKLLKREGTPPPPPPSRDLTEAYKTQALGPLKLKPAHEGLVATVKEAG

401 RSIHEIPREELRHTPELPLAPRPLKEGSITQGTPLKYDTGASTTGSKKHD

451 VRSLIGSPGRTFPPVHPLDVMADARALERACYEESLKSRPGTASSSGGSI

501 ARGAPVIVPELGKPRQSPLTYEDHGAPFAGHLPRGSPVTMREPTPRLQEG

551 SLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHPISPYEHLLRGVSGVD

601 LYRSHIPLAFDPTSIPRGIPLDAAAAYYLPRHLAPNPTYPHLYPPYLIRG

651 YPDTAALENRQTIINDYITSQQMHHNTATAMAQRADMLRGLSPRESSLAL

701 NYAAGPRGIIDLSQVPHLPVLVPPTPGTPATAMDRLAYLPTAPQPFSSRH

751 SSSPLSPGGPTHLTKPTTTSSSERERDRDRERDRDREREKSILTSTTTVE

801 HAPIWRPGTEQSSGSSGSSGGGGGSSSRPASHSHAHQHSPISPRTQDALQ

851 QRPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPVRPAATFPPATHCPLG

901 GTLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFLAKPPARSGLEPA

→ C-SMRT

951 SSPSKGSEPRPLVPPVSGHATIARTPAKNLAPHHASPDPPAPPASASDPH

1001 REKTQSKPFSIQELELRSLGYHGSSYSPEGVEPVSPVSSPSLTHDKGLPK

1051 HLEELDKSHLEGELRPKQPGPVKLGGEAAHLPHLRPLPESQPSSSPLIQT

1101 APGVKGHQRVVTLAQHISEVITQDYTRHHPQQLSAPLPAPLYSFPGASCP

1151 VLDLRRPPSDLYLPPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGI

1201 EPVSPPEGMTEPGHSRSAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFF

1251 SKLTESNSAMVKSKKQEINKKLNTHNRNEPEYNISQPGTEIFNMPAITGT

1301 GLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDQW.EESPPLSANAFNPL

1350 NASASLPAAMPITAADGRSDHTLTSP.GGGGKAKVSGRPSSRKAKSPAPG

1399 LA..SGDRPPSVSSVHSEGDCNRRTPLTNRVWEDRPSSAGSTPFPYNPLI

1447 MRLQAGVMASPPPPGLPAGSGPL..AGPHHA...WDEEPKPLLCSQYETL

FIGURE 2

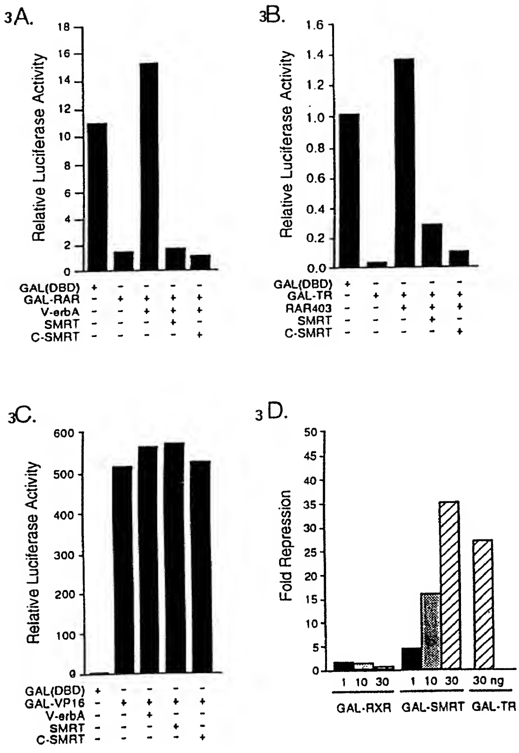
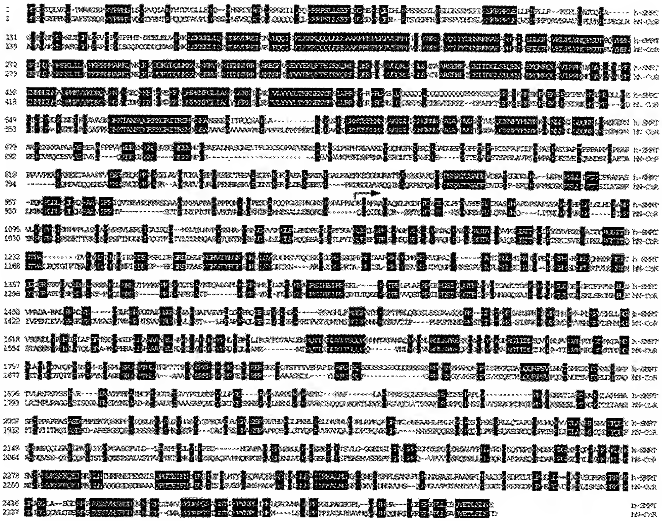

3A.
Relative Luciferase Activity
0
2
4
6
8
10
12
14
16
18
GAL(DBD) + - - - -
GAL-RAR - + + + +
V-erbA - - + + +
SMRT - - - + -
C-SMRT - - - - +
3B.
Relative Luciferase Activity
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
GAL(DBD) + - - - -
GAL-TR - + + + +
RAR403 - - + + +
SMRT - - - + -
C-SMRT - - - - +
3C.
Relative Luciferase Activity
0
100
200
300
400
500
600
GAL(DBD) + - - - -
GAL-VP16 - + + + +
V-erbA - - + - -
SMRT - - - + -
C-SMRT - - - - +
3 D.
Fold Repression
0
5
10
15
20
25
30
35
40
45
50
1 10 30
1 10 30
30 ng
GAL-RXR
GAL-SMRT
GAL-TR

FIGURE 3

FIGURE 4

5 A



5 B

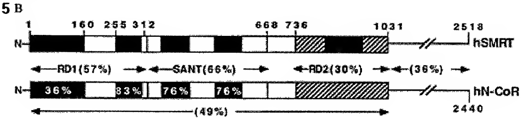


FIGURE 5

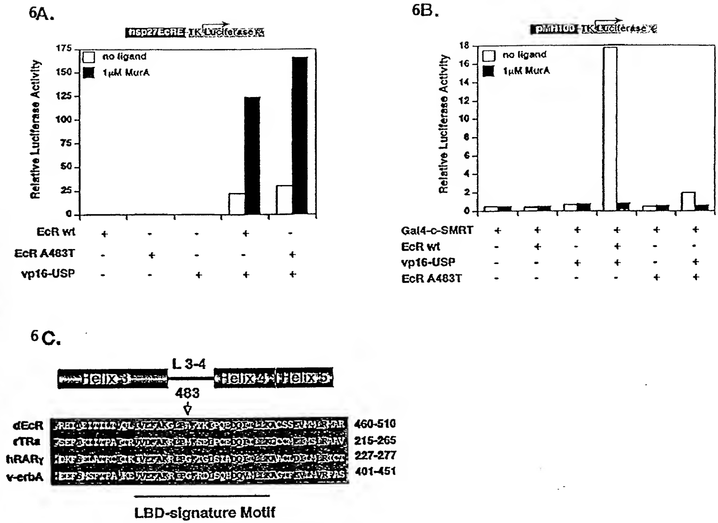
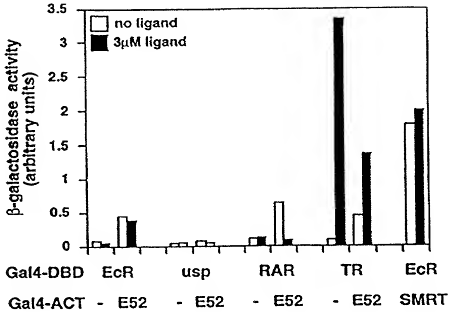

6A.
hsp27EcRE TK Luciferase
no ligand
1μM MurA
Relative Luciferase Activity
EcR wt
EcR A483T
vp16-USP
6B.
pMH100 TK Luciferase
no ligand
1μM MurA
Relative Luciferase Activity
Gal4-c-SMRT
EcR wt
vp16-USP
EcR A483T
6C.
Helix 3
L 3-4
Helix 4
Helix 5
483
dEcR
rTRα
hRARγ
v-erbA
460-510
215-265
227-277
401-451
LBD-signature Motif

FIGURE 6



3.5
3
2.5
2
1.5
1
0.5
0
no ligand
3μM ligand
β-galactosidase activity
(arbitrary units)
Gal4-DBD
EcR
usp
RAR
TR
EcR
Gal4-ACT
-
E52
-
E52
-
E52
-
E52
SMRT

FIGURE 7

DECLARATION FOR PATENT APPLICATION

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship is as stated below next to my name.

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled **A FAMILY OF TRANSCRIPTIONAL CO-REPRESSORS THAT INTERACT WITH NUCLEAR HORMONE RECEPTORS AND USES THEREFOR**, which is a C-I-P of 08/522,726, filed on September 1, 1995, the specification of which

__X__ is attached hereto. (SALK1510-3)

__X__ was filed on March 10, 2000, as U.S. Application Serial No. ____________________, and was amended on ____________________, if applicable (the "Application").

I hereby authorize and request insertion of the application serial number of the Application when officially known.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability of the subject matter of the Application as defined in Title 37, Code of Federal Regulations ("C.F.R."), § 1.56.

With respect to the Application, I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

(Application Serial No.)	(Filing Date)
(Application Serial No.)	(Filing Date)
(Application Serial No.)	(Filing Date)

With respect to the Application, I hereby claim the benefit under 35 U.S.C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of the application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability of the subject matter of the Application as defined in Title 37, C.F.R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of the Application:

08/522,726 (Application Serial No.)	09/01/95 (Filing Date)	pending (Status) (patented, pending, abandoned)
(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)
(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of first inventor: Ronald M. Evans

Inventor's signature: ____________________

Date: ____________________

Residence: La Jolla, California

Citizenship: USA

Post Office Address: 1471 Cottontail Lane
La Jolla, California 92037

Full name of second inventor: J. Don Chen

Inventor's signature: ____________________

Date: ____________________

Residence: San Diego, California

Citizenship: Taiwan

Post Office Address: 7548 Charmant Drive, #1416
San Diego, California 92126

Full name of third inventor: Peter Ordentlich

Inventor's signature: ____________________

Date: ____________________

Residence: ____________________

Citizenship: ____________________

Post Office Address: ____________________
____

8A.

1 MSAYQQRLPSNAASIHSPHWSYRALEQQQQYAKQAAHLQQQQHQSHQQQQQQQQDQRTNLHLQIHHHHQQQQQQQQQQQQ 80

81 QQQQQQQQQQQKQQQHHMQQQQQQQPLSPPHPPGSSSNSSSAAAAAAAAAAAAAAVNPGYPPSSAAAAAVNSGYPPRPPQ 160

161 HRFIQNTGYSIAPAPTYRDNPYSRHTQIQQQQQQQQQQQQQQQQQQQQQQQQAAASMPEYQRAAARAAVAAVSAGKGNVS 240

241 GQSSNSSSSSGGGGGGGSAGGSAPPGGGVVQVSQSGGVLVMEAMPHYASQPNSNPSQQQQQQQQQQQQGGNPSGAGATSG 320

321 AGGGGGGSGGSVMVGSLGRILMPHPQALQYTSEYLTNATAAVAAAMVNQRQHLQLQQQQQQQHPPEPFGGQQPYKKQRLS 400

401 EANANNMNHLPPHPQQQHQQQQQQQQQHQRSSPAQVQQQQQQQMNSSRQSHNDMCRQVVTTPMGMQLKVETLPQQQQKQQ 480

481 QHQQQQQQQQQGRSQPVVSSMSTVVSQPVGTVTVTTAGLSASHSGSSGNVAAGLGTGNTGSASTEAYHPQVEAISPTLPS 560

561 DSSIEERGRTSAKEDLLMQIQKVDNEIKSAETTMETLRKKEKSLMEEAALAKEQRAAKELNDNNNDQEPMVELSWRSQML 640

641 AEKIYAANRKTAQAQHSMLQNAAADESSPGSVAGRPWLPLYNQPLDVEALAMLIRQHQSQIRAPLLLHIRKLKAERWAHN 720

721 QGLVEKYTKDQADWQRRCERMEASAKRKAREAKNREFFEKVFTELRKQREDKERFNRVGSRIKSEADLEEIMDGLQEQAL 800

801 EDKKMRSYAVIPPLMHDARQRRCAYHNENGLIEDMVAVHQQRKALNMWTAGEKETFKEKYLQHPKNFGAIAASLDRKSPQ 880

881 DCVRYYYLSKKTENYKQLLRKSRQRTRSSRNPAKAQAAQPQCIIDSMTTGVMTRLQREQQQKSGGRSSAVAERERAERAA 960

961 ERERVAEKAAADAAKAAESAAEKASAATKAVEATAAGEKVAKAAAAAAAAAATTATTATTTTSSSTSSSSSSASSASTAS 1040

1041 SSTASPATLAGIAADKTDAGKTASASDKNAATAGGPTATGTPTAATTPATATAPPEISAGGEAKSKNAEEEAAATAGAAT 1120

1121 VATAGTPATGASAASAGEATTATGATATAAAKGVGKPETATEPAGTAAKGADSRPDANDPLAKTASKAINAEGYNAIGGN 1200

1201 SSSSSSNATGASAPVQGVTLNGFKPGYQTVVMANVKASTGGDDSGANAGGAAPGSLAATNASIATSGDKIVKTTPSSRAP 1280

1281 NSTSSTAANESSSGAGVNTYGHTATTAGNYLGQKLKAAQVEGLGAGNELHSDVSESKRKRFELNSGEAGGNATSAMTNSS 1360

1361 TSGSMNISNSHGLKANAKDGSMMAKTSMASTSSASVVVTSTPSASSSSLSSASSMLLISAASVMSTAAGATSSSTATTTA 1440

1441 TASAISLPLLADGSGNSMVNANEILALDGKDKLASCFVCKAEACPRTRPLKKGRGQQYGIPDETIPAGARVCNSCQCKSV 1520

1521 RSRYPNCPLPTCPNPKDRAQRLRNIPSRLFELAPEVRDPLMAEFQIPPHATRCCSACLMRIRRKLDPQLNLTDGSSGGAG 1600

1601 SGSGGDETDVSTSSCDEREPGGSDTASVESPENLQRHKSLTMVKQQQQQQQQQQQQQQQQQQQQQQQLSQPQPPPPAPQQ 1680

1681 QKGSSGRGGDQGTPLIITPTRMSSKSGSGGAQTAGDNERLLPPAAGQAPKKQKTSEEYDSSATETADEENENSPANRQSP 1760

1761 KVLFHGHGHGHGGHANNVAGLQPPVANMGTGGGVQPGGAAGQQVNGPISMRREAVNNVQDCVFSVIERSLKHKGPQPKGG 1840

1841 QGQQQGQGQGQGQGQGQTPGQSQSPSQQQQQQQQQQSANNLERKELTIVREYRQDPGILKQQQQQQQAGGAPPTSAAGSL 1920

1921 PHGTSVQKLTTRPAAVAPPPPAHPLTPTSIGCAGSNNGTSDSLATLSVVNSHMGNVGIGHPGPMAHASSAGGIGVDKATI 2000

2001 TPVVKSSSGSSKSGGGSASSHSTATPPETIIYNVPVAHPQRGIPPPSQHSVHPAHPSHTQHPAHPQHSSHGQHTQLQVPE 2080

2081 PEPQTLDLSIKKPPRDGHSPHTGAGGSSSSGSGSGGPSSSDRHHGPPPPTMSMKHIVRSGGMYRGDTVTVPSLAAPSSYL 2160

2161 YPTRSVKSIGGGGVVPGVLPGVPGIALYLQPVPVPVPISISGQGQLPPKAGQPPPAQPPSGRGVAKVPPKLSPQQAHHLH 2240

2241 PSHGHSPSQQQQQQQQQQQQQQQQAAAAQQQLLVKSGSIIHGTPANSAQQQIIVHAPATAAAAPSSLFSPKFDGLVRQTT 2320

2321 PEGVGSVGPGGASGSGKHGSITQGTPLHMPPHHLESKRPYESYYKSSQRHSPAQQPGGNQQLPPPPQQSSPQAPPPQGYG 2400

2401 VGVSSPYARSPFAGVVEQPQVLSTRQIVMHDYITSQQMQGQQQQQQQQQQQQRNHSRGSSASGGGGGGGSDKESPSPRNS 2480

2481 VGSASGFAYGGDKESAPRGRPEYSSRASPADHVNSTPSPHRTPPPQRQGVIQRHNTGSKPPSPAAPPPSRMHMPPYQYAP 2560

2561 SGHDALASFVDVAVQQPQLPVPSQKDDKSPGPSTAPGQVPGSGPPLGPSPLPPHAVVGVAQPPPPTAHHDQRYRDLTLHH 2640

2641 HHHTLVQQQIAQQQHYRSLNVAAQVDMQRQMDQAKRVMRHQQHQVQQQQQQQQQQQHNHALERDREMQERMRERDRERER 2720

2721 EREREQRERERERERERERERERERREQDRARRVVAEEREHDSRRMERMFAGNVVTGSGGAGGGGPSPGQFLRASVPETGPP 2800

2801 RSIPDRERESYYRQAHGGPAPEDTPGQLSAQSLIDAIIKHEINRSNDATAGPGREFPRPSFVHAPLPPRGSGSGGGTGTR 2880

2881 SSPANVLHPHYLRDLRQPLDGGAGSMLTAENNGKPSSSGSPSVINIDLDQERISAAAAAVAQQQQQQQAPPPSQSSQSRS 2960

2961 VHGQLRTPTSQSGGSAPSPQQIHTKSITFGELTDSIITSDYGTNPHLRPPYMAYLQETQSIIPPDRMKQNRRMQQKAEEA 3040

3041 NDHSQQQQQQQHQQQHHAQQQQQQQQQQHHAQQHHPQMPGTGSGSAPGGAGQGGGSGG[illegible]GGGGAGRASTPGEDGRNLTR 3120

3121 KPQAVSPRKFNHEMMLHHVNGTTGAGGEAGQFFLPSRVVLPEQRGTPSGGGGAPGAGGPGSGGGATTIEKYVATRIALVM 3200

3201 KDDIGYGKNRTVEVRTEDEVTADKVAHSHAAVHAAHVAHAAHVAHAAAMLLQHRSKEPPPPP[illegible]ISVSRKTPHQYKVVDASG 3280

3281 RRSAGSGSVSYSVSGANSHHSPYHPPAAAYAPSTYAPPYSALNYPGAAGGLPPHQPLQIAHQAVAPPGAPAKAKAAHALS 3360

3361 ELGAVGGGVSLVVGGGSGGIAGGPGGVSVGVGVPGGGGPGSGGGGGGGHNSSSSQASAAVAAAVAAAASESKPLLLSKYD 3440

3441 ALSDED*

8B.

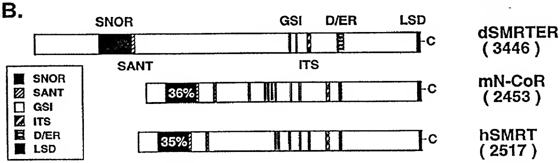


FIGURE 8

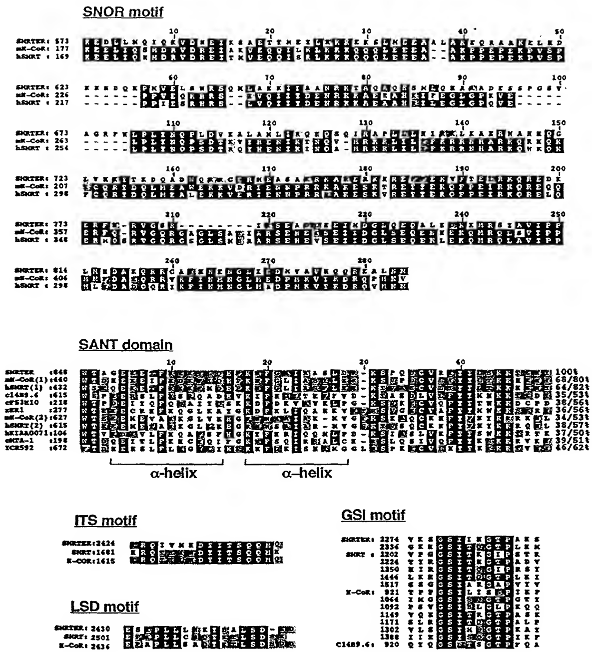

SNOR motif
SANT domain
α-helix
α–helix
ITS motif
GSI motif
LSD motif

FIGURE 9

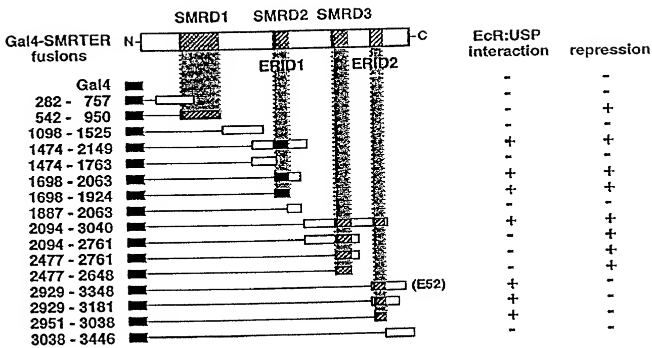
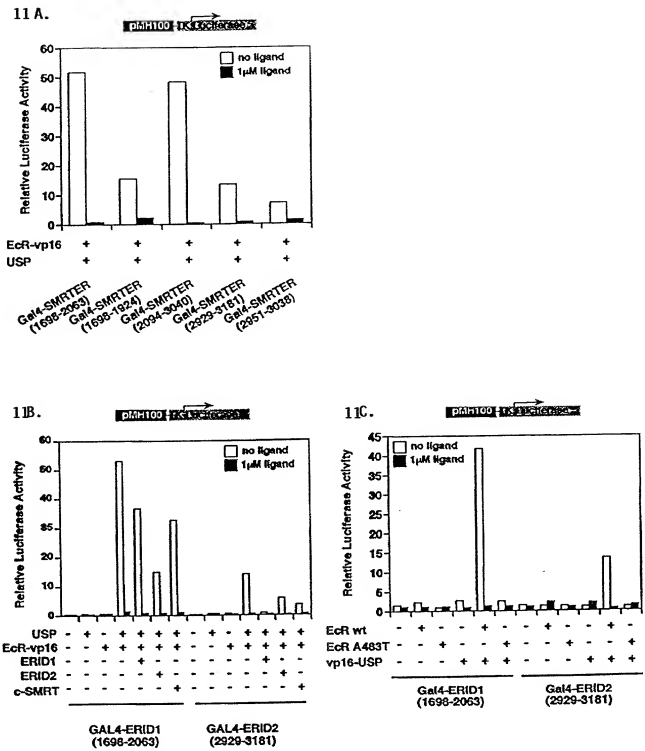

SMRD1
SMRD2
SMRD3
Gal4-SMRTER fusions
N
C
ERID1
ERID2
EcR:USP interaction
repression
Gal4
282 - 757
542 - 950
1098 - 1525
1474 - 2149
1474 - 1763
1698 - 2063
1698 - 1924
1887 - 2063
2094 - 3040
2094 - 2761
2477 - 2761
2477 - 2648
2929 - 3348
(E52)
2929 - 3181
2951 - 3038
3038 - 3446

FIGURE 10



11A.
pMH100 TK Luciferase
no ligand
1μM ligand
Relative Luciferase Activity
EcR-vp16
USP
Gal4-SMRTER (1698-2063)
Gal4-SMRTER (1698-1824)
Gal4-SMRTER (2094-3040)
Gal4-SMRTER (2929-3181)
Gal4-SMRTER (2951-3038)
11B.
pMH100 TK Luciferase
no ligand
1μM ligand
Relative Luciferase Activity
USP
EcR-vp16
ERID1
ERID2
c-SMRT
GAL4-ERID1 (1698-2063)
GAL4-ERID2 (2929-3181)
11C.
pMH100 TK Luciferase
no ligand
1μM ligand
Relative Luciferase Activity
EcR wt
EcR A483T
vp16-USP
Gal4-ERID1 (1698-2063)
Gal4-ERID2 (2929-3181)

FIGURE 11

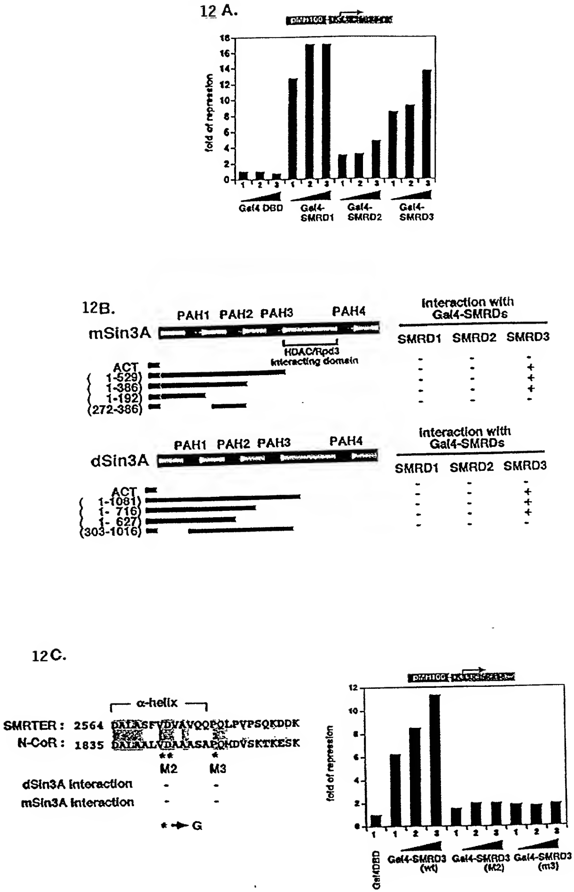

12 A.
fold of repression
Gal4 DBD
Gal4-SMRD1
Gal4-SMRD2
Gal4-SMRD3
12B.
PAH1 PAH2 PAH3 PAH4
mSin3A
HDAC/Rpd3 interacting domain
ACT
1-529
1-386
1-192
272-386
Interaction with Gal4-SMRDs
SMRD1 SMRD2 SMRD3
dSin3A
1-1081
1- 716
1- 627
303-1016
12C.
α-helix
SMRTER : 2564 DALASFVDVAVQQPQLPVPSQKDDK
N-CoR : 1835 DALAALVDAAASAPQHDVSKTKESK
M2 M3
dSin3A interaction
mSin3A interaction
* → G
Gal4DBD
Gal4-SMRD3 (wt)
Gal4-SMRD3 (M2)
Gal4-SMRD3 (m3)

FIGURE 12

SEQUENCE LISTING

<110> Evans, Ronald M.
Chen, J. Don

<120> A FAMILY OF TRANSCRIPTIONAL
CO-REPRESSORS THAT INTERACT WITH NUCLEAR HORMONE RECEPTORS
AND USES THEREFOR

<130> SALK1510-3

<150> 09/337,384
<151> 1999-06-21

<150> 08/522,726
<151> 1995-09-01

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1495
<212> PRT
<213> Homo sapiens

<400> 1
Met Glu Ala Trp Asp Ala His Pro Asp Lys Glu Ala Phe Ala Ala Glu
1 5 10 15
Ala Gln Lys Leu Pro Gly Asp Pro Pro Cys Trp Thr Ser Gly Leu Pro
20 25 30
Phe Pro Val Pro Pro Arg Glu Val Ile Lys Ala Ser Pro His Ala Pro
35 40 45
Asp Pro Ser Ala Phe Ser Tyr Ala Pro Pro Gly His Pro Leu Pro Leu
50 55 60
Gly Leu His Asp Thr Ala Arg Pro Val Leu Pro Arg Pro Pro Thr Ile
65 70 75 80
Ser Asn Pro Pro Pro Leu Ile Ser Ser Ala Lys His Pro Ser Val Leu
85 90 95
Glu Arg Gln Ile Gly Ala Ile Ser Gln Gly Met Ser Val Gln Leu His
100 105 110
Val Pro Tyr Ser Glu His Ala Lys Ala Pro Val Gly Pro Val Thr Met
115 120 125
Gly Leu Pro Leu Pro Met Asp Pro Lys Lys Leu Ala Pro Phe Ser Gly
130 135 140
Val Lys Gln Glu Gln Leu Ser Pro Arg Gly Gln Ala Gly Pro Pro Glu
145 150 155 160
Ser Leu Gly Val Pro Thr Ala Gln Glu Ala Ser Val Leu Arg Gly Thr
165 170 175
Ala Leu Gly Ser Val Pro Gly Gly Ser Ile Thr Lys Gly Ile Pro Ser
180 185 190
Thr Arg Val Pro Ser Asp Ser Ala Ile Thr Tyr Arg Gly Ser Ile Thr
195 200 205
His Gly Thr Pro Ala Asp Val Leu Tyr Lys Gly Thr Ile Thr Arg Ile
210 215 220
Ile Gly Glu Asp Ser Pro Ser Arg Leu Asp Arg Gly Arg Glu Asp Ser
225 230 235 240
Leu Pro Lys Gly His Val Ile Tyr Glu Gly Lys Lys Gly His Val Leu
245 250 255

Ser Tyr Glu Gly Gly Met Ser Val Thr Gln Cys Ser Lys Glu Asp Gly
260 265 270
Arg Ser Ser Ser Gly Pro Pro His Glu Thr Ala Ala Pro Lys Arg Thr
275 280 285
Tyr Asp Met Met Glu Gly Arg Val Gly Arg Ala Ile Ser Ser Ala Ser
290 295 300
Ile Glu Gly Leu Met Gly Arg Ala Ile Pro Pro Glu Arg His Ser Pro
305 310 315 320
His His Leu Lys Glu Gln His His Ile Arg Gly Ser Ile Thr Gln Gly
325 330 335
Ile Pro Arg Ser Tyr Val Glu Ala Gln Glu Asp Tyr Leu Arg Arg Glu
340 345 350
Ala Lys Leu Leu Lys Arg Glu Gly Thr Pro Pro Pro Pro Pro Pro Ser
355 360 365
Arg Asp Leu Thr Glu Ala Tyr Lys Thr Gln Ala Leu Gly Pro Leu Lys
370 375 380
Leu Lys Pro Ala His Glu Gly Leu Val Ala Thr Val Lys Glu Ala Gly
385 390 395 400
Arg Ser Ile His Glu Ile Pro Arg Glu Glu Leu Arg His Thr Pro Glu
405 410 415
Leu Pro Leu Ala Pro Arg Pro Leu Lys Glu Gly Ser Ile Thr Gln Gly
420 425 430
Thr Pro Leu Lys Tyr Asp Thr Gly Ala Ser Thr Thr Gly Ser Lys Lys
435 440 445
His Asp Val Arg Ser Leu Ile Gly Ser Pro Gly Arg Thr Phe Pro Pro
450 455 460
Val His Pro Leu Asp Val Met Ala Asp Ala Arg Ala Leu Glu Arg Ala
465 470 475 480
Cys Tyr Glu Glu Ser Leu Lys Ser Arg Pro Gly Thr Ala Ser Ser Ser
485 490 495
Gly Gly Ser Ile Ala Arg Gly Ala Pro Val Ile Val Pro Glu Leu Gly
500 505 510
Lys Pro Arg Gln Ser Pro Leu Thr Tyr Glu Asp His Gly Ala Pro Phe
515 520 525
Ala Gly His Leu Pro Arg Gly Ser Pro Val Thr Met Arg Glu Pro Thr
530 535 540
Pro Arg Leu Gln Glu Gly Ser Leu Ser Ser Ser Lys Ala Ser Gln Asp
545 550 555 560
Arg Lys Leu Thr Ser Thr Pro Arg Glu Ile Ala Lys Ser Pro His Ser
565 570 575
Thr Val Pro Glu His His Pro His Pro Ile Ser Pro Tyr Glu His Leu
580 585 590
Leu Arg Gly Val Ser Gly Val Asp Leu Tyr Arg Ser His Ile Pro Leu
595 600 605
Ala Phe Asp Pro Thr Ser Ile Pro Arg Gly Ile Pro Leu Asp Ala Ala
610 615 620
Ala Ala Tyr Tyr Leu Pro Arg His Leu Ala Pro Asn Pro Thr Tyr Pro
625 630 635 640
His Leu Tyr Pro Pro Tyr Leu Ile Arg Gly Tyr Pro Asp Thr Ala Ala
645 650 655
Leu Glu Asn Arg Gln Thr Ile Ile Asn Asp Tyr Ile Thr Ser Gln Gln
660 665 670
Met His His Asn Thr Ala Thr Ala Met Ala Gln Arg Ala Asp Met Leu
675 680 685
Arg Gly Leu Ser Pro Arg Glu Ser Ser Leu Ala Leu Asn Tyr Ala Ala
690 695 700
Gly Pro Arg Gly Ile Ile Asp Leu Ser Gln Val Pro His Leu Pro Val
705 710 715 720
Leu Val Pro Pro Thr Pro Gly Thr Pro Ala Thr Ala Met Asp Arg Leu
725 730 735

Ala Tyr Leu Pro Thr Ala Pro Gln Pro Phe Ser Ser Arg His Ser Ser
740 745 750
Ser Pro Leu Ser Pro Gly Gly Pro Thr His Leu Thr Lys Pro Thr Thr
755 760 765
Thr Ser Ser Ser Glu Arg Glu Arg Asp Arg Asp Arg Glu Arg Asp Arg
770 775 780
Asp Arg Glu Arg Glu Lys Ser Ile Leu Thr Ser Thr Thr Thr Val Glu
785 790 795 800
His Ala Pro Ile Trp Arg Pro Gly Thr Glu Gln Ser Ser Gly Ser Ser
805 810 815
Gly Ser Ser Gly Gly Gly Gly Gly Ser Ser Ser Arg Pro Ala Ser His
820 825 830
Ser His Ala His Gln His Ser Pro Ile Ser Pro Arg Thr Gln Asp Ala
835 840 845
Leu Gln Gln Arg Pro Ser Val Leu His Asn Thr Gly Met Lys Gly Ile
850 855 860
Ile Thr Ala Val Glu Pro Ser Lys Pro Thr Val Leu Arg Ser Thr Ser
865 870 875 880
Thr Ser Ser Pro Val Arg Pro Ala Ala Thr Phe Pro Pro Ala Thr His
885 890 895
Cys Pro Leu Gly Gly Thr Leu Asp Gly Val Tyr Pro Thr Leu Met Glu
900 905 910
Pro Val Leu Leu Pro Lys Glu Ala Pro Arg Val Ala Arg Pro Glu Arg
915 920 925
Pro Arg Ala Asp Thr Gly His Ala Phe Leu Ala Lys Pro Pro Ala Arg
930 935 940
Ser Gly Leu Glu Pro Ala Ser Ser Pro Ser Lys Gly Ser Glu Pro Arg
945 950 955 960
Pro Leu Val Pro Pro Val Ser Gly His Ala Thr Ile Ala Arg Thr Pro
965 970 975
Ala Lys Asn Leu Ala Pro His His Ala Ser Pro Asp Pro Pro Ala Pro
980 985 990
Pro Ala Ser Ala Ser Asp Pro His Arg Glu Lys Thr Gln Ser Lys Pro
995 1000 1005
Phe Ser Ile Gln Glu Leu Glu Leu Arg Ser Leu Gly Tyr His Gly Ser
1010 1015 1020
Ser Tyr Ser Pro Glu Gly Val Glu Pro Val Ser Pro Val Ser Ser Pro
1025 1030 1035 1040
Ser Leu Thr His Asp Lys Gly Leu Pro Lys His Leu Glu Glu Leu Asp
1045 1050 1055
Lys Ser His Leu Glu Gly Glu Leu Arg Pro Lys Gln Pro Gly Pro Val
1060 1065 1070
Lys Leu Gly Gly Glu Ala Ala His Leu Pro His Leu Arg Pro Leu Pro
1075 1080 1085
Glu Ser Gln Pro Ser Ser Ser Pro Leu Leu Gln Thr Ala Pro Gly Val
1090 1095 1100
Lys Gly His Gln Arg Val Val Thr Leu Ala Gln His Ile Ser Glu Val
1105 1110 1115 1120
Ile Thr Gln Asp Tyr Thr Arg His His Pro Gln Gln Leu Ser Ala Pro
1125 1130 1135
Leu Pro Ala Pro Leu Tyr Ser Phe Pro Gly Ala Ser Cys Pro Val Leu
1140 1145 1150
Asp Leu Arg Arg Pro Pro Ser Asp Leu Tyr Leu Pro Pro Pro Asp His
1155 1160 1165
Gly Ala Pro Ala Arg Gly Ser Pro His Ser Glu Gly Gly Lys Arg Ser
1170 1175 1180
Pro Glu Pro Asn Lys Thr Ser Val Leu Gly Gly Gly Glu Asp Gly Ile
1185 1190 1195 1200
Glu Pro Val Ser Pro Pro Glu Gly Met Thr Glu Pro Gly His Ser Arg
1205 1210 1215

Ser Ala Val Tyr Pro Leu Leu Tyr Arg Asp Gly Glu Gln Thr Glu Pro
1220 1225 1230
Ser Arg Met Gly Ser Lys Ser Pro Gly Asn Thr Ser Gln Pro Pro Ala
1235 1240 1245
Phe Phe Ser Lys Leu Thr Glu Ser Asn Ser Ala Met Val Lys Ser Lys
1250 1255 1260
Lys Gln Glu Ile Asn Lys Lys Leu Asn Thr His Asn Arg Asn Glu Pro
1265 1270 1275 1280
Glu Tyr Asn Ile Ser Gln Pro Gly Thr Glu Ile Phe Asn Met Pro Ala
1285 1290 1295
Ile Thr Gly Thr Gly Leu Met Thr Tyr Arg Ser Gln Ala Val Gln Glu
1300 1305 1310
His Ala Ser Thr Asn Met Gly Leu Glu Ala Ile Ile Arg Lys Ala Leu
1315 1320 1325
Met Gly Lys Tyr Asp Gln Trp Glu Glu Ser Pro Pro Leu Ser Ala Asn
1330 1335 1340
Ala Phe Asn Pro Leu Asn Ala Ser Ala Ser Leu Pro Ala Ala Met Pro
1345 1350 1355 1360
Ile Thr Ala Ala Asp Gly Arg Ser Asp His Thr Leu Thr Ser Pro Gly
1365 1370 1375
Gly Gly Gly Lys Ala Lys Val Ser Gly Arg Pro Ser Ser Arg Lys Ala
1380 1385 1390
Lys Ser Pro Ala Pro Gly Leu Ala Ser Gly Asp Arg Pro Pro Ser Val
1395 1400 1405
Ser Ser Val His Ser Glu Gly Asp Cys Asn Arg Arg Thr Pro Leu Thr
1410 1415 1420
Asn Arg Val Trp Glu Asp Arg Pro Ser Ser Ala Gly Ser Thr Pro Phe
1425 1430 1435 1440
Pro Tyr Asn Pro Leu Ile Met Arg Leu Gln Ala Gly Tyr Met Ala Ser
1445 1450 1455
Pro Pro Pro Pro Gly Leu Pro Ala Gly Ser Gly Pro Leu Ala Gly Pro
1460 1465 1470
His His Ala Trp Asp Glu Glu Pro Lys Pro Leu Leu Cys Ser Gln Tyr
1475 1480 1485
Glu Thr Leu Ser Asp Ser Glu
1490 1495

<210> 2
<211> 46
<212> PRT
<213> Homo sapiens

<400> 2
Gly Lys Tyr Asp Gln Trp Glu Glu Ser Pro Pro Leu Ser Ala Asn Ala
1 5 10 15
Phe Asn Pro Leu Asn Ala Ser Ala Ser Leu Pro Ala Ala Met Pro Ile
20 25 30
Thr Ala Ala Asp Gly Arg Ser Asp His Thr Leu Thr Ser Pro
35 40 45

<210> 3
<211> 17
<212> DNA
<213> Homo sapiens

<400> 3
cggaggactg tcctccg 17

<210> 4
<211> 8561

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (2)...(7555)

<400> 4
c atg tcg ggc tcc aca cag ctt gtg gca cag acg tgg agg gcc act gag 49
Met Ser Gly Ser Thr Gln Leu Val Ala Gln Thr Trp Arg Ala Thr Glu
1 5 10 15

ccc cgc tac ccg ccc cac agc ctt tcc tac cca gtg cag atc gcc cgg 97
Pro Arg Tyr Pro Pro His Ser Leu Ser Tyr Pro Val Gln Ile Ala Arg
20 25 30

acg cac acg gac gtc ggg ctc ctg gag tac cag cac cac tcc cgc gac 145
Thr His Thr Asp Val Gly Leu Leu Glu Tyr Gln His His Ser Arg Asp
35 40 45

tat gcc tcc cac ctg tcg ccg ggc tcc atc atc cag ccc cag cgg cgg 193
Tyr Ala Ser His Leu Ser Pro Gly Ser Ile Ile Gln Pro Gln Arg Arg
50 55 60

agg ccc tcc ctg ctg tct gag ttc cag ccc ggg aat gaa cgg tcc cag 241
Arg Pro Ser Leu Leu Ser Glu Phe Gln Pro Gly Asn Glu Arg Ser Gln
65 70 75 80

gag ctc cac ctg cgg cca gag tcc cac tca tac ctg ccc gag ctg ggg 289
Glu Leu His Leu Arg Pro Glu Ser His Ser Tyr Leu Pro Glu Leu Gly
85 90 95

aag tca gag atg gag ttc att gaa agc aag cgc cct cgg cta gag ctg 337
Lys Ser Glu Met Glu Phe Ile Glu Ser Lys Arg Pro Arg Leu Glu Leu
100 105 110

ctg cct gac ccc ctg ctg cga ccg tca ccc ctg ctg gcc acg ggc cag 385
Leu Pro Asp Pro Leu Leu Arg Pro Ser Pro Leu Leu Ala Thr Gly Gln
115 120 125

cct gcg gga tct gaa gac ctc acc aag gac cgt agc ctg acg ggc aag 433
Pro Ala Gly Ser Glu Asp Leu Thr Lys Asp Arg Ser Leu Thr Gly Lys
130 135 140

ctg gaa ccg gtg tct ccc ccc agc ccc ccg cac act gac cct gag ctg 481
Leu Glu Pro Val Ser Pro Pro Ser Pro Pro His Thr Asp Pro Glu Leu
145 150 155 160

gag ctg gtg ccg cca cgg ctg tcc aag gag gag ctg atc cag aac atg 529
Glu Leu Val Pro Pro Arg Leu Ser Lys Glu Glu Leu Ile Gln Asn Met
165 170 175

gac cgc gtg gac cga gag atc acc atg gta gag cag cag atc tct aag 577
Asp Arg Val Asp Arg Glu Ile Thr Met Val Glu Gln Gln Ile Ser Lys
180 185 190

ctg aag aag aag cag caa cag ctg gag gag gag gct gcc aag ccg ccc 625
Leu Lys Lys Lys Gln Gln Gln Leu Glu Glu Glu Ala Ala Lys Pro Pro
195 200 205

```
gag cct gag aag ccc gtg tca ccg ccg ccc atc gag tcg aag cac cgc       673
Glu Pro Glu Lys Pro Val Ser Pro Pro Pro Ile Glu Ser Lys His Arg
    210                 215                 220

agc ctg gtg cag atc atc tac gac gag aac cgg aag aag gct gaa gct       721
Ser Leu Val Gln Ile Ile Tyr Asp Glu Asn Arg Lys Lys Ala Glu Ala
225                 230                 235                 240

gca cat cgg att ctg gaa ggc ctg ggg ccc cag gtg gag ctg ccg ctg       769
Ala His Arg Ile Leu Glu Gly Leu Gly Pro Gln Val Glu Leu Pro Leu
                245                 250                 255

tac aac cag ccc tcc gac acc cgg cag tat cat gag aac atc aaa ata       817
Tyr Asn Gln Pro Ser Asp Thr Arg Gln Tyr His Glu Asn Ile Lys Ile
            260                 265                 270

aac cag gcg atg cgg aag aag cta atc ttg tac ttc aag agg agg aat       865
Asn Gln Ala Met Arg Lys Lys Leu Ile Leu Tyr Phe Lys Arg Arg Asn
        275                 280                 285

cac gct cgg aaa caa tgg aag cag aag ttc tgc cag cgc tat gac cag       913
His Ala Arg Lys Gln Trp Lys Gln Lys Phe Cys Gln Arg Tyr Asp Gln
    290                 295                 300

ctc atg gag gcc ttg gaa aaa aag gtg gag cgc atc gaa aac aac ccg       961
Leu Met Glu Ala Leu Glu Lys Lys Val Glu Arg Ile Glu Asn Asn Pro
305                 310                 315                 320

cgc cgg cgg gcc aag gag agc aag gtg cgc gag tac tac gaa aag cag      1009
Arg Arg Arg Ala Lys Glu Ser Lys Val Arg Glu Tyr Tyr Glu Lys Gln
                325                 330                 335

ttc cct gag atc cgc aag cag cgc gag ctg cag gag cgc atg cag agc      1057
Phe Pro Glu Ile Arg Lys Gln Arg Glu Leu Gln Glu Arg Met Gln Ser
            340                 345                 350

agg gtg ggc cag cgg ggc agt ggg ctg tcc atg tcg gcc gcc cgc agc      1105
Arg Val Gly Gln Arg Gly Ser Gly Leu Ser Met Ser Ala Ala Arg Ser
        355                 360                 365

gag cac gag gtg tca gag atc atc gat ggc ctc tca gag cag gag aac      1153
Glu His Glu Val Ser Glu Ile Ile Asp Gly Leu Ser Glu Gln Glu Asn
    370                 375                 380

ctg gag aag cag atg cgc cag ctg gcc gtg atc ccg ccc atg ctg tac      1201
Leu Glu Lys Gln Met Arg Gln Leu Ala Val Ile Pro Pro Met Leu Tyr
385                 390                 395                 400

gac gct gac cag cag cgc atc aag ttc atc aac atg aac ggg ctt atg      1249
Asp Ala Asp Gln Gln Arg Ile Lys Phe Ile Asn Met Asn Gly Leu Met
                405                 410                 415

gcc gac ccc atg aag gtg tac aaa gac cgc cag gtc atg aac atg tgg      1297
Ala Asp Pro Met Lys Val Tyr Lys Asp Arg Gln Val Met Asn Met Trp
            420                 425                 430

agt gag cag gag aag gag acc ttc cgg gag aag ttc atg cag cat ccc      1345
Ser Glu Gln Glu Lys Glu Thr Phe Arg Glu Lys Phe Met Gln His Pro
        435                 440                 445
```

```
aag aac ttt ggc ctg atc gca tca ttc ctg gag agg aag aca gtg gct      1393
Lys Asn Phe Gly Leu Ile Ala Ser Phe Leu Glu Arg Lys Thr Val Ala
    450                 455                 460

gag tgc gtc ctc tat tac tac ctg act aag aag aat gag aac tat aag      1441
Glu Cys Val Leu Tyr Tyr Tyr Leu Thr Lys Lys Asn Glu Asn Tyr Lys
465                 470                 475                 480

agc ctg gtg aga cgg agc tat cgg cgc cgc ggc aag agc cag cag caa      1489
Ser Leu Val Arg Arg Ser Tyr Arg Arg Arg Gly Lys Ser Gln Gln Gln
                485                 490                 495

caa cag cag cag cag cag cag cag cag cag cag cag cag cag ccc atg      1537
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Met
            500                 505                 510

ccc cgc agc agc cag gag gag aaa gat gag aag gag aag gaa aag gag      1585
Pro Arg Ser Ser Gln Glu Glu Lys Asp Glu Lys Glu Lys Glu Lys Glu
        515                 520                 525

gcg gag aag gag gag gag aag ccg gag gtg gag aac gac aag gaa gac      1633
Ala Glu Lys Glu Glu Glu Lys Pro Glu Val Glu Asn Asp Lys Glu Asp
    530                 535                 540

ctc ctc aag gag aag aca gac gac acc tca ggg gag gac aac gac gag      1681
Leu Leu Lys Glu Lys Thr Asp Asp Thr Ser Gly Glu Asp Asn Asp Glu
545                 550                 555                 560

aag gag gct gtg gcc tcc aaa ggc cgc aaa act gcc aac agc cag gga      1729
Lys Glu Ala Val Ala Ser Lys Gly Arg Lys Thr Ala Asn Ser Gln Gly
                565                 570                 575

aga cgc aaa ggc cgc atc acc cgc tca atg gct aat gag gcc aac agc      1777
Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Ala Asn Glu Ala Asn Ser
            580                 585                 590

gag gag gcc atc acc ccc cag cag agc gcc gag ctg gcc tcc atg gag      1825
Glu Glu Ala Ile Thr Pro Gln Gln Ser Ala Glu Leu Ala Ser Met Glu
        595                 600                 605

ctg aat gag agt tct cgc tgg aca gaa gaa gaa atg gaa aca gcc aag      1873
Leu Asn Glu Ser Ser Arg Trp Thr Glu Glu Glu Met Glu Thr Ala Lys
    610                 615                 620

aaa ggt ctc ctg gaa cac ggc cgc aac tgg tcg gcc atc gcc cgg atg      1921
Lys Gly Leu Leu Glu His Gly Arg Asn Trp Ser Ala Ile Ala Arg Met
625                 630                 635                 640

gtg ggc tcc aag act gtg tcg cag tgt aag aac ttc tac ttc aac tac      1969
Val Gly Ser Lys Thr Val Ser Gln Cys Lys Asn Phe Tyr Phe Asn Tyr
                645                 650                 655

aag aag agg cag aac ctc gat gag atc ttg cag cag cac aag ctg aag      2017
Lys Lys Arg Gln Asn Leu Asp Glu Ile Leu Gln Gln His Lys Leu Lys
            660                 665                 670

atg gag aag gag agg aac gcg cgg agg aag aag aag aaa gcg ccg gcg      2065
Met Glu Lys Glu Arg Asn Ala Arg Arg Lys Lys Lys Lys Ala Pro Ala
        675                 680                 685
```

gcg gcc agc gag gag gct gca ttc ccg ccc gtg gtg gag gat gag gag 2113
Ala Ala Ser Glu Glu Ala Ala Phe Pro Pro Val Val Glu Asp Glu Glu
690 695 700

atg gag gcg tcg ggc gtg agc gga aat gag gag gag atg gtg gag gag 2161
Met Glu Ala Ser Gly Val Ser Gly Asn Glu Glu Glu Met Val Glu Glu
705 710 715 720

gct gaa gcc tta cat gcc tct ggg aat gag gtg ccc aga ggg gaa tgc 2209
Ala Glu Ala Leu His Ala Ser Gly Asn Glu Val Pro Arg Gly Glu Cys
725 730 735

agt ggc cca gcc act gtc aac aac agc tca gac acc gag agc atc ccc 2257
Ser Gly Pro Ala Thr Val Asn Asn Ser Ser Asp Thr Glu Ser Ile Pro
740 745 750

tct cct cac act gag gcc gcc aag gac aca ggg cag aat ggg ccc aag 2305
Ser Pro His Thr Glu Ala Ala Lys Asp Thr Gly Gln Asn Gly Pro Lys
755 760 765

ccc cca gcc acc ctg ggc gcc gac ggg cca ccc cca ggc cca ccc acc 2353
Pro Pro Ala Thr Leu Gly Ala Asp Gly Pro Pro Pro Gly Pro Pro Thr
770 775 780

cca cca cgg agg aca tcc cgg gcc ccc att gag ccc acc ccg gcc tct 2401
Pro Pro Arg Arg Thr Ser Arg Ala Pro Ile Glu Pro Thr Pro Ala Ser
785 790 795 800

gaa gcc acc gga gcc cct acg ccc cca cca gca ccc cca tcg ccc tct 2449
Glu Ala Thr Gly Ala Pro Thr Pro Pro Pro Ala Pro Pro Ser Pro Ser
805 810 815

gca cct cct cct gtg gtc ccc aag gag gag aag gag gag gag acc gca 2497
Ala Pro Pro Pro Val Val Pro Lys Glu Glu Lys Glu Glu Glu Thr Ala
820 825 830

gca gcg ccc cca gtg gag gag ggg gag gag cag aag ccc ccc gcg gct 2545
Ala Ala Pro Pro Val Glu Glu Gly Glu Glu Gln Lys Pro Pro Ala Ala
835 840 845

gag gag ctg gca gtg gac aca ggg aag gcc gag gag ccc gtc aag agc 2593
Glu Glu Leu Ala Val Asp Thr Gly Lys Ala Glu Glu Pro Val Lys Ser
850 855 860

gag tgc acg gag gaa gcc gag gag ggg ccg gcc aag ggc aag gac gcg 2641
Glu Cys Thr Glu Glu Ala Glu Glu Gly Pro Ala Lys Gly Lys Asp Ala
865 870 875 880

gag gcc gct gag gcc acg gcc gag ggg gcg ctc aag gca gag aag aag 2689
Glu Ala Ala Glu Ala Thr Ala Glu Gly Ala Leu Lys Ala Glu Lys Lys
885 890 895

gag ggc ggg agc ggc agg gcc acc act gcc aag agc tcg ggc gcc ccc 2737
Glu Gly Gly Ser Gly Arg Ala Thr Thr Ala Lys Ser Ser Gly Ala Pro
900 905 910

cag gac agc gac tcc agt gct acc tgc agt gca gac gag gtg gat gag 2785
Gln Asp Ser Asp Ser Ser Ala Thr Cys Ser Ala Asp Glu Val Asp Glu
915 920 925

gcc gag ggc ggc gac aag aac cgg ctg ctg tcc cca agg ccc agc ctc 2833
Ala Glu Gly Gly Asp Lys Asn Arg Leu Leu Ser Pro Arg Pro Ser Leu
930 935 940

ctc acc ccg act ggc gac ccc cgg gcc aat gcc tca ccc cag aag cca 2881
Leu Thr Pro Thr Gly Asp Pro Arg Ala Asn Ala Ser Pro Gln Lys Pro
945 950 955 960

ctg gac ctg aag cag ctg aag cag cga gcg gct gcc atc ccc ccc atc 2929
Leu Asp Leu Lys Gln Leu Lys Gln Arg Ala Ala Ala Ile Pro Pro Ile
965 970 975

cag gtc acc aaa gtc cat gag ccc ccc cgg gag gac gca gct ccc acc 2977
Gln Val Thr Lys Val His Glu Pro Pro Arg Glu Asp Ala Ala Pro Thr
980 985 990

aag cca gct ccc cca gcc cca ccg cca ccg caa aac ctg cag ccg gag 3025
Lys Pro Ala Pro Pro Ala Pro Pro Pro Pro Gln Asn Leu Gln Pro Glu
995 1000 1005

agc gac gcc cct cag cag cct ggc agc agc ccc cgg ggc aag agc agg 3073
Ser Asp Ala Pro Gln Gln Pro Gly Ser Ser Pro Arg Gly Lys Ser Arg
1010 1015 1020

agc ccg gca ccc ccc gcc gac aag gag gcc ttc gca gcc gag gcc cag 3121
Ser Pro Ala Pro Pro Ala Asp Lys Glu Ala Phe Ala Ala Glu Ala Gln
1025 1030 1035 1040

aag ctg cct ggg gac ccc cct tgc tgg act tcc ggc ctg ccc ttc ccc 3169
Lys Leu Pro Gly Asp Pro Pro Cys Trp Thr Ser Gly Leu Pro Phe Pro
1045 1050 1055

gtg ccc ccc cgt gag gtg atc aag gcc tcc ccg cat gcc ccg gac ccc 3217
Val Pro Pro Arg Glu Val Ile Lys Ala Ser Pro His Ala Pro Asp Pro
1060 1065 1070

tca gcc ttc tcc tac gct cca cct ggt cac cca ctg ccc ctg ggc ctc 3265
Ser Ala Phe Ser Tyr Ala Pro Pro Gly His Pro Leu Pro Leu Gly Leu
1075 1080 1085

cat gac act gcc cgg ccc gtc ctg ccg cgc cca ccc acc atc tcc aac 3313
His Asp Thr Ala Arg Pro Val Leu Pro Arg Pro Pro Thr Ile Ser Asn
1090 1095 1100

ccg cct ccc ctc atc tcc tct gcc aag cac ccc agc gtc ctc gag agg 3361
Pro Pro Pro Leu Ile Ser Ser Ala Lys His Pro Ser Val Leu Glu Arg
1105 1110 1115 1120

caa ata ggt gcc atc tcc caa gga atg tcg gtc cag ctc cac gtc ccg 3409
Gln Ile Gly Ala Ile Ser Gln Gly Met Ser Val Gln Leu His Val Pro
1125 1130 1135

tac tca gag cat gcc aag gcc ccg gtg ggc cct gtc acc atg ggg ctg 3457
Tyr Ser Glu His Ala Lys Ala Pro Val Gly Pro Val Thr Met Gly Leu
1140 1145 1150

ccc ctg ccc atg gac ccc aaa aag ctg gca ccc ttc agc gga gtg aag 3505
Pro Leu Pro Met Asp Pro Lys Lys Leu Ala Pro Phe Ser Gly Val Lys
1155 1160 1165

cag gag cag ctg tcc cca cgg ggc cag gct ggg cca ccg gag agc ctg 3553
Gln Glu Gln Leu Ser Pro Arg Gly Gln Ala Gly Pro Pro Glu Ser Leu
1170 1175 1180

ggg gtg ccc aca gcc cag gag gcg tcc gtg ctg aga ggg aca gct ctg 3601
Gly Val Pro Thr Ala Gln Glu Ala Ser Val Leu Arg Gly Thr Ala Leu
1185 1190 1195 1200

ggc tca gtt ccg ggc gga agc atc acc aaa ggc att ccc agc aca cgg 3649
Gly Ser Val Pro Gly Gly Ser Ile Thr Lys Gly Ile Pro Ser Thr Arg
1205 1210 1215

gtg ccc tcg gac agc gcc atc aca tac cgc ggc tcc atc acc cac ggc 3697
Val Pro Ser Asp Ser Ala Ile Thr Tyr Arg Gly Ser Ile Thr His Gly
1220 1225 1230

acg cca gct gac gtc ctg tac aag ggc acc atc acc agg atc atc ggc 3745
Thr Pro Ala Asp Val Leu Tyr Lys Gly Thr Ile Thr Arg Ile Ile Gly
1235 1240 1245

gag gac agc ccg agt cgc ttg gac cgc ggc cgg gag gac agc ctg ccc 3793
Glu Asp Ser Pro Ser Arg Leu Asp Arg Gly Arg Glu Asp Ser Leu Pro
1250 1255 1260

aag ggc cac gtc atc tac gaa ggc aag aag ggc cac gtc ttg tcc tat 3841
Lys Gly His Val Ile Tyr Glu Gly Lys Lys Gly His Val Leu Ser Tyr
1265 1270 1275 1280

gag ggt ggc atg tct gtg acc cag tgc tcc aag gag gac ggc aga agc 3889
Glu Gly Gly Met Ser Val Thr Gln Cys Ser Lys Glu Asp Gly Arg Ser
1285 1290 1295

agc tca gga ccc ccc cat gag acg gcc gcc ccc aag cgc acc tat gac 3937
Ser Ser Gly Pro Pro His Glu Thr Ala Ala Pro Lys Arg Thr Tyr Asp
1300 1305 1310

atg atg gag ggc cgc gtg ggc aga gcc atc tcc tca gcc agc atc gaa 3985
Met Met Glu Gly Arg Val Gly Arg Ala Ile Ser Ser Ala Ser Ile Glu
1315 1320 1325

ggt ctc atg ggc cgt gcc atc ccg ccg gag cga cac agc ccc cac cac 4033
Gly Leu Met Gly Arg Ala Ile Pro Pro Glu Arg His Ser Pro His His
1330 1335 1340

ctc aaa gag cag cac cac atc cgc ggg tcc atc aca caa ggg atc cct 4081
Leu Lys Glu Gln His His Ile Arg Gly Ser Ile Thr Gln Gly Ile Pro
1345 1350 1355 1360

cgg tcc tac gtg gag gca cag gag gac tac ctg cgt cgg gag gcc aag 4129
Arg Ser Tyr Val Glu Ala Gln Glu Asp Tyr Leu Arg Arg Glu Ala Lys
1365 1370 1375

ctc cta aag cgg gag ggc acg cct ccg ccc cca ccg ccc tca cgg gac 4177
Leu Leu Lys Arg Glu Gly Thr Pro Pro Pro Pro Pro Pro Ser Arg Asp
1380 1385 1390

ctg acc gag gcc tac aag acg cag gcc ctg ggc ccc ctg aag ctg aag 4225
Leu Thr Glu Ala Tyr Lys Thr Gln Ala Leu Gly Pro Leu Lys Leu Lys
1395 1400 1405

ccg gcc cat gag ggc ctg gtg gcc acg gtg aag gag gcg ggc cgc tcc 4273
Pro Ala His Glu Gly Leu Val Ala Thr Val Lys Glu Ala Gly Arg Ser
1410 1415 1420

atc cat gag atc ccg cgc gag gag ctg cgg cac acg ccc gag ctg ccc 4321
Ile His Glu Ile Pro Arg Glu Glu Leu Arg His Thr Pro Glu Leu Pro
1425 1430 1435 1440

ctg gcc ccg cgg ccg ctc aag gag ggc tcc atc acg cag ggc acc ccg 4369
Leu Ala Pro Arg Pro Leu Lys Glu Gly Ser Ile Thr Gln Gly Thr Pro
1445 1450 1455

ctc aag tac gac acc ggc gcg tcc acc act ggc tcc aaa aag cac gac 4417
Leu Lys Tyr Asp Thr Gly Ala Ser Thr Thr Gly Ser Lys Lys His Asp
1460 1465 1470

gta cgc tcc ctc atc ggc agc ccc ggc cgg acg ttc cca ccc gtg cac 4465
Val Arg Ser Leu Ile Gly Ser Pro Gly Arg Thr Phe Pro Pro Val His
1475 1480 1485

ccg ctg gat gtg atg gcc gac gcc cgg gca ctg gaa cgt gcc tgc tac 4513
Pro Leu Asp Val Met Ala Asp Ala Arg Ala Leu Glu Arg Ala Cys Tyr
1490 1495 1500

gag gag agc ctg aag agc cgg cca ggg acc gcc agc agc tcg ggg ggc 4561
Glu Glu Ser Leu Lys Ser Arg Pro Gly Thr Ala Ser Ser Ser Gly Gly
1505 1510 1515 1520

tcc att gcg cgc ggc gcc ccg gtc att gtg cct gag ctg ggt aag ccg 4609
Ser Ile Ala Arg Gly Ala Pro Val Ile Val Pro Glu Leu Gly Lys Pro
1525 1530 1535

cgg cag agc ccc ctg acc tat gag gac cac ggg gca ccc ttt gcc ggc 4657
Arg Gln Ser Pro Leu Thr Tyr Glu Asp His Gly Ala Pro Phe Ala Gly
1540 1545 1550

cac ctc cca cga ggt tcg ccc gtg acc atg cgg gag ccc acg ccg cgc 4705
His Leu Pro Arg Gly Ser Pro Val Thr Met Arg Glu Pro Thr Pro Arg
1555 1560 1565

ctg cag gag ggc agc ctt tcg tcc agc aag gca tcc cag gac cga aag 4753
Leu Gln Glu Gly Ser Leu Ser Ser Ser Lys Ala Ser Gln Asp Arg Lys
1570 1575 1580

ctg acg tcg acg cct cgt gag atc gcc aag tcc ccg cac agc acc gtg 4801
Leu Thr Ser Thr Pro Arg Glu Ile Ala Lys Ser Pro His Ser Thr Val
1585 1590 1595 1600

ccc gag cac cac cca cac ccc atc tcg ccc tat gag cac ctg ctt cgg 4849
Pro Glu His His Pro His Pro Ile Ser Pro Tyr Glu His Leu Leu Arg
1605 1610 1615

ggc gtg agt ggc gtg gac ctg tat cgc agc cac atc ccc ctg gcc ttc 4897
Gly Val Ser Gly Val Asp Leu Tyr Arg Ser His Ile Pro Leu Ala Phe
1620 1625 1630

gac ccc acc tcc ata ccc cgc ggc atc cct ctg gac gca gcc gct gcc 4945
Asp Pro Thr Ser Ile Pro Arg Gly Ile Pro Leu Asp Ala Ala Ala Ala
1635 1640 1645

tac tac ctg ccc cga cac ctg gcc ccc aac ccc acc tac ccg cac ctg 4993
Tyr Tyr Leu Pro Arg His Leu Ala Pro Asn Pro Thr Tyr Pro His Leu
1650 1655 1660

tac cca ccc tac ctc atc cgc ggc tac ccc gac acg gcg gcg ctg gag 5041
Tyr Pro Pro Tyr Leu Ile Arg Gly Tyr Pro Asp Thr Ala Ala Leu Glu
1665 1670 1675 1680

aac cgg cag acc atc atc aat gac tac atc acc tcg cag cag atg cac 5089
Asn Arg Gln Thr Ile Ile Asn Asp Tyr Ile Thr Ser Gln Gln Met His
1685 1690 1695

cac aac acg gcc acc gcc atg gcc cag cga gct gat atg ctg agg ggc 5137
His Asn Thr Ala Thr Ala Met Ala Gln Arg Ala Asp Met Leu Arg Gly
1700 1705 1710

ctc tcg ccc cgc gag tcc tcg ctg gca ctc aac tac gct gcg ggt ccc 5185
Leu Ser Pro Arg Glu Ser Ser Leu Ala Leu Asn Tyr Ala Ala Gly Pro
1715 1720 1725

cga ggc atc atc gac ctg tcc caa gtg cca cac ctg cct gtg ctc gtg 5233
Arg Gly Ile Ile Asp Leu Ser Gln Val Pro His Leu Pro Val Leu Val
1730 1735 1740

ccc ccg aca cca ggc acc cca gcc acc gcc atg gac cgc ctt gcc tac 5281
Pro Pro Thr Pro Gly Thr Pro Ala Thr Ala Met Asp Arg Leu Ala Tyr
1745 1750 1755 1760

ctc ccc acc gcg ccc cag ccc ttc agc agc cgc cac agc agc tcc cca 5329
Leu Pro Thr Ala Pro Gln Pro Phe Ser Ser Arg His Ser Ser Ser Pro
1765 1770 1775

ctc tcc cca gga ggt cca aca cac ttg aca aaa cca acc acc acg tcc 5377
Leu Ser Pro Gly Gly Pro Thr His Leu Thr Lys Pro Thr Thr Thr Ser
1780 1785 1790

tcg tcc gag cgg gag cga gac cgg gat cga gag cgg gac cgg gat cgg 5425
Ser Ser Glu Arg Glu Arg Asp Arg Asp Arg Glu Arg Asp Arg Asp Arg
1795 1800 1805

gag cgg gaa aag tcc atc ctc acg tcc acc acg acg gtg gag cac gca 5473
Glu Arg Glu Lys Ser Ile Leu Thr Ser Thr Thr Thr Val Glu His Ala
1810 1815 1820

ccc atc tgg aga cct ggt aca gag cag agc agc ggc agc agc ggc agc 5521
Pro Ile Trp Arg Pro Gly Thr Glu Gln Ser Ser Gly Ser Ser Gly Ser
1825 1830 1835 1840

agc ggc ggg ggt ggg ggc agc agc agc cgc ccc gcc tcc cac tcc cat 5569
Ser Gly Gly Gly Gly Gly Ser Ser Ser Arg Pro Ala Ser His Ser His
1845 1850 1855

gcc cac cag cac tcg ccc atc tcc cct cgg acc cag gat gcc ctc cag 5617
Ala His Gln His Ser Pro Ile Ser Pro Arg Thr Gln Asp Ala Leu Gln
1860 1865 1870

cag aga ccc agt gtg ctt cac aac aca ggc atg aag ggt atc atc acc 5665
Gln Arg Pro Ser Val Leu His Asn Thr Gly Met Lys Gly Ile Ile Thr
1875 1880 1885

gct gtg gag ccc agc aag ccc acg gtc ctg agg tcc acc tcc acc tcc 5713
Ala Val Glu Pro Ser Lys Pro Thr Val Leu Arg Ser Thr Ser Thr Ser
1890 1895 1900

tca ccc gtt cgc cca gct gcc aca ttc cca cct gcc acc cac tgc cca 5761
Ser Pro Val Arg Pro Ala Ala Thr Phe Pro Pro Ala Thr His Cys Pro
1905 1910 1915 1920

ctg ggc ggc acc ctc gat ggg gtc tac cct acc ctc atg gag ccc gtc 5809
Leu Gly Gly Thr Leu Asp Gly Val Tyr Pro Thr Leu Met Glu Pro Val
1925 1930 1935

ttg ctg ccc aag gag gcc ccc cgg gtc gcc cgg cca gag cgg ccc cga 5857
Leu Leu Pro Lys Glu Ala Pro Arg Val Ala Arg Pro Glu Arg Pro Arg
1940 1945 1950

gca gac acc ggc cat gcc ttc ctc gcc aag ccc cca gcc cgc tcc ggg 5905
Ala Asp Thr Gly His Ala Phe Leu Ala Lys Pro Pro Ala Arg Ser Gly
1955 1960 1965

ctg gag ccc gcc tcc tcc ccc agc aag ggc tcg gag ccc cgg ccc cta 5953
Leu Glu Pro Ala Ser Ser Pro Ser Lys Gly Ser Glu Pro Arg Pro Leu
1970 1975 1980

gtg cct cct gtc tct ggc cac gcc acc atc gcc cgc acc cct gcg aag 6001
Val Pro Pro Val Ser Gly His Ala Thr Ile Ala Arg Thr Pro Ala Lys
1985 1990 1995 2000

aac ctc gca cct cac cac gcc agc ccg gac ccg ccg gcg cca cct gcc 6049
Asn Leu Ala Pro His His Ala Ser Pro Asp Pro Pro Ala Pro Pro Ala
2005 2010 2015

tcg gcc tcg gac ccg cac cgg gaa aag act caa agt aaa ccc ttt tcc 6097
Ser Ala Ser Asp Pro His Arg Glu Lys Thr Gln Ser Lys Pro Phe Ser
2020 2025 2030

atc cag gaa ctg gaa ctc cgt tct ctg ggt tac cac ggc agc agc tac 6145
Ile Gln Glu Leu Glu Leu Arg Ser Leu Gly Tyr His Gly Ser Ser Tyr
2035 2040 2045

agc ccc gaa ggg gtg gag ccc gtc agc cct gtg agc tca ccc agt ctg 6193
Ser Pro Glu Gly Val Glu Pro Val Ser Pro Val Ser Ser Pro Ser Leu
2050 2055 2060

acc cac gac aag ggg ctc ccc aag cac ctg gaa gag ctc gac aag agc 6241
Thr His Asp Lys Gly Leu Pro Lys His Leu Glu Glu Leu Asp Lys Ser
2065 2070 2075 2080

cac ctg gag ggg gag ctg cgg ccc aag cag cca ggc ccc gtg aag ctt 6289
His Leu Glu Gly Glu Leu Arg Pro Lys Gln Pro Gly Pro Val Lys Leu
2085 2090 2095

ggc ggg gag gcc gcc cac ctc cca cac ctg cgg ccg ctg cct gag agc 6337
Gly Gly Glu Ala Ala His Leu Pro His Leu Arg Pro Leu Pro Glu Ser
2100 2105 2110

cag ccc tcg tcc agc ccg ctg ctc cag acc gcc cca ggg gtc aaa ggt 6385
Gln Pro Ser Ser Ser Pro Leu Leu Gln Thr Ala Pro Gly Val Lys Gly
2115 2120 2125

```
cac cag cgg gtg gtc acc ctg gcc cag cac atc agt gag gtc atc aca      6433
His Gln Arg Val Val Thr Leu Ala Gln His Ile Ser Glu Val Ile Thr
    2130                2135                2140

cag gac tac acc cgg cac cac cca cag cag ctc agc gca ccc ctg ccc      6481
Gln Asp Tyr Thr Arg His His Pro Gln Gln Leu Ser Ala Pro Leu Pro
2145                2150                2155                2160

gcc ccc ctc tac tcc ttc cct ggg gcc agc tgc ccc gtc ctg gac ctc      6529
Ala Pro Leu Tyr Ser Phe Pro Gly Ala Ser Cys Pro Val Leu Asp Leu
                2165                2170                2175

cgc cgc cca ccc agt gac ctc tac ctc ccg ccc ccg gac cat ggt gcc      6577
Arg Arg Pro Pro Ser Asp Leu Tyr Leu Pro Pro Pro Asp His Gly Ala
            2180                2185                2190

ccg gcc cgt ggc tcc ccc cac agc gaa ggg ggc aag agg tct cca gag      6625
Pro Ala Arg Gly Ser Pro His Ser Glu Gly Gly Lys Arg Ser Pro Glu
        2195                2200                2205

cca aac aag acg tcg gtc ttg ggt ggt ggt gag gac ggt att gaa cct      6673
Pro Asn Lys Thr Ser Val Leu Gly Gly Gly Glu Asp Gly Ile Glu Pro
    2210                2215                2220

gtg tcc cca ccg gag ggc atg acg gag cca ggg cac tcc cgg agt gct      6721
Val Ser Pro Pro Glu Gly Met Thr Glu Pro Gly His Ser Arg Ser Ala
2225                2230                2235                2240

gtg tac ccg ctg ctg tac cgg gat ggg gaa cag acg gag ccc agc agg      6769
Val Tyr Pro Leu Leu Tyr Arg Asp Gly Glu Gln Thr Glu Pro Ser Arg
                2245                2250                2255

atg ggc tcc aag tct cca ggc aac acc agc cag ccg cca gcc ttc ttc      6817
Met Gly Ser Lys Ser Pro Gly Asn Thr Ser Gln Pro Pro Ala Phe Phe
            2260                2265                2270

agc aag ctg acc gag agc aac tcc gcc atg gtc aag tcc aag aag caa      6865
Ser Lys Leu Thr Glu Ser Asn Ser Ala Met Val Lys Ser Lys Lys Gln
        2275                2280                2285

gag atc aac aag aag ctg aac acc cac aac cgg aat gag cct gaa tac      6913
Glu Ile Asn Lys Lys Leu Asn Thr His Asn Arg Asn Glu Pro Glu Tyr
    2290                2295                2300

aat atc agc cag cct ggg acg gag atc ttc aat atg ccc gcc atc acc      6961
Asn Ile Ser Gln Pro Gly Thr Glu Ile Phe Asn Met Pro Ala Ile Thr
2305                2310                2315                2320

gga aca ggc ctt atg acc tat aga agc cag gcg gtg cag gaa cat gcc      7009
Gly Thr Gly Leu Met Thr Tyr Arg Ser Gln Ala Val Gln Glu His Ala
                2325                2330                2335

agc acc aac atg ggg ctg gag gcc ata att aga aag gca ctc atg ggt      7057
Ser Thr Asn Met Gly Leu Glu Ala Ile Ile Arg Lys Ala Leu Met Gly
            2340                2345                2350

aaa tat gac cag tgg gaa gag tcc ccg ccg ctc agc gcc aat gct ttt      7105
Lys Tyr Asp Gln Trp Glu Glu Ser Pro Pro Leu Ser Ala Asn Ala Phe
        2355                2360                2365
```

aac cct ctg aat gcc agt gcc agc ctg ccc gct gct atg ccc ata acc 7153
Asn Pro Leu Asn Ala Ser Ala Ser Leu Pro Ala Ala Met Pro Ile Thr
2370 2375 2380

gct gct gac gga cgg agt gac cac aca ctc acc tcg cca ggt ggc ggc 7201
Ala Ala Asp Gly Arg Ser Asp His Thr Leu Thr Ser Pro Gly Gly Gly
2385 2390 2395 2400

ggg aag gcc aag gtc tct ggc aga ccc agc agc cga aaa gcc aag tcc 7249
Gly Lys Ala Lys Val Ser Gly Arg Pro Ser Ser Arg Lys Ala Lys Ser
2405 2410 2415

ccg gcc ccg ggc ctg gca tct ggg gac cgg cca ccc tct gtc tcc tca 7297
Pro Ala Pro Gly Leu Ala Ser Gly Asp Arg Pro Pro Ser Val Ser Ser
2420 2425 2430

gtg cac tcg gag gga gac tgc aac cgc cgg acg ccg ctc acc aac cgc 7345
Val His Ser Glu Gly Asp Cys Asn Arg Arg Thr Pro Leu Thr Asn Arg
2435 2440 2445

gtg tgg gag gac agg ccc tcg tcc gca ggt tcc acg cca ttc ccc tac 7393
Val Trp Glu Asp Arg Pro Ser Ser Ala Gly Ser Thr Pro Phe Pro Tyr
2450 2455 2460

aac ccc ctg atc atg cgg ctg cag gcg ggt gtc atg gct tcc cca ccc 7441
Asn Pro Leu Ile Met Arg Leu Gln Ala Gly Val Met Ala Ser Pro Pro
2465 2470 2475 2480

cca ccg ggc ctc ccc gcg ggc agc ggg ccc ctc gct ggc ccc cac cac 7489
Pro Pro Gly Leu Pro Ala Gly Ser Gly Pro Leu Ala Gly Pro His His
2485 2490 2495

gcc tgg gac gag gag ccc aag cca ctg ctc tgc tcg cag tac gag aca 7537
Ala Trp Asp Glu Glu Pro Lys Pro Leu Leu Cys Ser Gln Tyr Glu Thr
2500 2505 2510

ctc tcc gac agc gag tga ctcagaacag ggcgggggg ggcgggcggt 7585
Leu Ser Asp Ser Glu *
2515

gtcaggtccc agcgagccac aggaacggcc ctgcaggagc ggggcggctg ccgactcccc 7645
caaccaagga aggagcccct gagtccgcct gcgcctccat ccatctgtcc gtccagagcc 7705
ggcatccttg cctgtctaaa gccttaacta agactcccgc cccgggctgg ccctgtgcag 7765
accttactca ggggatgttt acctggtgct cgggaaggga ggggaagggg ccggggaggg 7825
ggcacggcag gcgtgtggca gccacacaca ggcggccagg gcggccaggg acccaaagca 7885
ggatgaccac gcacctccac gccactgcct cccccgaatg catttggaac caaagtctaa 7945
actgagctcg cagcccccgc gccctccctc cgcctcccat cccgcttagc gctctggaca 8005
gatggacgca ggccctgtcc agcccccagt gcgctcgttc cggtccccac agactgcccc 8065
agccaacgag attgctggaa accaagtcag gccaggtggg cggacaaaag ggccaggtgc 8125
ggcctggggg gaacggatgc tccgaggact ggactgtttt tttcacacat cgttgccgca 8185
gcggtgggaa ggaaaggcag atgtaaatga tgtgttggtt tacagggtat atttttgata 8245
ccttcaatga attaattcag atgttttacg caaggaagga cttacccagt attactgctg 8305
ctgtgctttt gatctctgct taccgttcaa gaggcgtgtg caggccgaca gtcggtgacc 8365
ccatcactcg caggaccaag gggcgggga ctgctcgtca cgccccgctg tgtcctccct 8425
ccctcccttc cttgggcaga atgaattcga tgcgtattct gtggccgcca tttgcgcagg 8485
gtggtggtat tctgtcattt acacacgtcg ttctaattaa aaagcgaatt atactccaaa 8545
aaaaaaaaaa aaaaaa 8561

<210> 5
<211> 2517

<212> PRT
<213> Homo sapiens

<400> 5
Met Ser Gly Ser Thr Gln Leu Val Ala Gln Thr Trp Arg Ala Thr Glu
1   5   10   15
Pro Arg Tyr Pro Pro His Ser Leu Ser Tyr Pro Val Gln Ile Ala Arg
            20   25   30
Thr His Thr Asp Val Gly Leu Leu Glu Tyr Gln His His Ser Arg Asp
        35   40   45
Tyr Ala Ser His Leu Ser Pro Gly Ser Ile Ile Gln Pro Gln Arg Arg
    50   55   60
Arg Pro Ser Leu Leu Ser Glu Phe Gln Pro Gly Asn Glu Arg Ser Gln
65   70   75   80
Glu Leu His Leu Arg Pro Glu Ser His Ser Tyr Leu Pro Glu Leu Gly
                85   90   95
Lys Ser Glu Met Glu Phe Ile Glu Ser Lys Arg Pro Arg Leu Glu Leu
            100   105   110
Leu Pro Asp Pro Leu Leu Arg Pro Ser Pro Leu Leu Ala Thr Gly Gln
        115   120   125
Pro Ala Gly Ser Glu Asp Leu Thr Lys Asp Arg Ser Leu Thr Gly Lys
    130   135   140
Leu Glu Pro Val Ser Pro Pro Ser Pro Pro His Thr Asp Pro Glu Leu
145   150   155   160
Glu Leu Val Pro Pro Arg Leu Ser Lys Glu Glu Leu Ile Gln Asn Met
                165   170   175
Asp Arg Val Asp Arg Glu Ile Thr Met Val Glu Gln Gln Ile Ser Lys
            180   185   190
Leu Lys Lys Lys Gln Gln Gln Leu Glu Glu Glu Ala Ala Lys Pro Pro
        195   200   205
Glu Pro Glu Lys Pro Val Ser Pro Pro Pro Ile Glu Ser Lys His Arg
    210   215   220
Ser Leu Val Gln Ile Ile Tyr Asp Glu Asn Arg Lys Lys Ala Glu Ala
225   230   235   240
Ala His Arg Ile Leu Glu Gly Leu Gly Pro Gln Val Glu Leu Pro Leu
                245   250   255
Tyr Asn Gln Pro Ser Asp Thr Arg Gln Tyr His Glu Asn Ile Lys Ile
            260   265   270
Asn Gln Ala Met Arg Lys Lys Leu Ile Leu Tyr Phe Lys Arg Arg Asn
        275   280   285
His Ala Arg Lys Gln Trp Lys Gln Lys Phe Cys Gln Arg Tyr Asp Gln
    290   295   300
Leu Met Glu Ala Leu Glu Lys Lys Val Glu Arg Ile Glu Asn Asn Pro
305   310   315   320
Arg Arg Arg Ala Lys Glu Ser Lys Val Arg Glu Tyr Tyr Glu Lys Gln
                325   330   335
Phe Pro Glu Ile Arg Lys Gln Arg Glu Leu Gln Glu Arg Met Gln Ser
            340   345   350
Arg Val Gly Gln Arg Gly Ser Gly Leu Ser Met Ser Ala Ala Arg Ser
        355   360   365
Glu His Glu Val Ser Glu Ile Ile Asp Gly Leu Ser Glu Gln Glu Asn
    370   375   380
Leu Glu Lys Gln Met Arg Gln Leu Ala Val Ile Pro Pro Met Leu Tyr
385   390   395   400
Asp Ala Asp Gln Gln Arg Ile Lys Phe Ile Asn Met Asn Gly Leu Met
                405   410   415
Ala Asp Pro Met Lys Val Tyr Lys Asp Arg Gln Val Met Asn Met Trp
            420   425   430
Ser Glu Gln Glu Lys Glu Thr Phe Arg Glu Lys Phe Met Gln His Pro
        435   440   445

Lys Asn Phe Gly Leu Ile Ala Ser Phe Leu Glu Arg Lys Thr Val Ala
450 455 460

Glu Cys Val Leu Tyr Tyr Tyr Leu Thr Lys Lys Asn Glu Asn Tyr Lys
465 470 475 480

Ser Leu Val Arg Arg Ser Tyr Arg Arg Arg Gly Lys Ser Gln Gln Gln
485 490 495

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Met
500 505 510

Pro Arg Ser Ser Gln Glu Glu Lys Asp Glu Lys Glu Lys Glu Lys Glu
515 520 525

Ala Glu Lys Glu Glu Glu Lys Pro Glu Val Glu Asn Asp Lys Glu Asp
530 535 540

Leu Leu Lys Glu Lys Thr Asp Asp Thr Ser Gly Glu Asp Asn Asp Glu
545 550 555 560

Lys Glu Ala Val Ala Ser Lys Gly Arg Lys Thr Ala Asn Ser Gln Gly
565 570 575

Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Ala Asn Glu Ala Asn Ser
580 585 590

Glu Glu Ala Ile Thr Pro Gln Gln Ser Ala Glu Leu Ala Ser Met Glu
595 600 605

Leu Asn Glu Ser Ser Arg Trp Thr Glu Glu Glu Met Glu Thr Ala Lys
610 615 620

Lys Gly Leu Leu Glu His Gly Arg Asn Trp Ser Ala Ile Ala Arg Met
625 630 635 640

Val Gly Ser Lys Thr Val Ser Gln Cys Lys Asn Phe Tyr Phe Asn Tyr
645 650 655

Lys Lys Arg Gln Asn Leu Asp Glu Ile Leu Gln Gln His Lys Leu Lys
660 665 670

Met Glu Lys Glu Arg Asn Ala Arg Arg Lys Lys Lys Lys Ala Pro Ala
675 680 685

Ala Ala Ser Glu Glu Ala Ala Phe Pro Pro Val Val Glu Asp Glu Glu
690 695 700

Met Glu Ala Ser Gly Val Ser Gly Asn Glu Glu Glu Met Val Glu Glu
705 710 715 720

Ala Glu Ala Leu His Ala Ser Gly Asn Glu Val Pro Arg Gly Glu Cys
725 730 735

Ser Gly Pro Ala Thr Val Asn Asn Ser Ser Asp Thr Glu Ser Ile Pro
740 745 750

Ser Pro His Thr Glu Ala Ala Lys Asp Thr Gly Gln Asn Gly Pro Lys
755 760 765

Pro Pro Ala Thr Leu Gly Ala Asp Gly Pro Pro Pro Gly Pro Pro Thr
770 775 780

Pro Pro Arg Arg Thr Ser Arg Ala Pro Ile Glu Pro Thr Pro Ala Ser
785 790 795 800

Glu Ala Thr Gly Ala Pro Thr Pro Pro Pro Ala Pro Pro Ser Pro Ser
805 810 815

Ala Pro Pro Pro Val Val Pro Lys Glu Glu Lys Glu Glu Glu Thr Ala
820 825 830

Ala Ala Pro Pro Val Glu Glu Gly Glu Glu Gln Lys Pro Pro Ala Ala
835 840 845

Glu Glu Leu Ala Val Asp Thr Gly Lys Ala Glu Glu Pro Val Lys Ser
850 855 860

Glu Cys Thr Glu Glu Ala Glu Glu Gly Pro Ala Lys Gly Lys Asp Ala
865 870 875 880

Glu Ala Ala Glu Ala Thr Ala Glu Gly Ala Leu Lys Ala Glu Lys Lys
885 890 895

Glu Gly Gly Ser Gly Arg Ala Thr Thr Ala Lys Ser Ser Gly Ala Pro
900 905 910

Gln Asp Ser Asp Ser Ser Ala Thr Cys Ser Ala Asp Glu Val Asp Glu
915 920 925

```
Ala Glu Gly Gly Asp Lys Asn Arg Leu Leu Ser Pro Arg Pro Ser Leu
    930                 935                 940
Leu Thr Pro Thr Gly Asp Pro Arg Ala Asn Ala Ser Pro Gln Lys Pro
945                 950                 955                 960
Leu Asp Leu Lys Gln Leu Lys Gln Arg Ala Ala Ala Ile Pro Pro Ile
                965                 970                 975
Gln Val Thr Lys Val His Glu Pro Pro Arg Glu Asp Ala Ala Pro Thr
            980                 985                 990
Lys Pro Ala Pro Pro Ala Pro Pro Pro Pro Gln Asn Leu Gln Pro Glu
        995                 1000                1005
Ser Asp Ala Pro Gln Gln Pro Gly Ser Ser Pro Arg Gly Lys Ser Arg
    1010                1015                1020
Ser Pro Ala Pro Pro Ala Asp Lys Glu Ala Phe Ala Ala Glu Ala Gln
1025                1030                1035                1040
Lys Leu Pro Gly Asp Pro Pro Cys Trp Thr Ser Gly Leu Pro Phe Pro
                1045                1050                1055
Val Pro Pro Arg Glu Val Ile Lys Ala Ser Pro His Ala Pro Asp Pro
            1060                1065                1070
Ser Ala Phe Ser Tyr Ala Pro Pro Gly His Pro Leu Pro Leu Gly Leu
        1075                1080                1085
His Asp Thr Ala Arg Pro Val Leu Pro Arg Pro Pro Thr Ile Ser Asn
    1090                1095                1100
Pro Pro Pro Leu Ile Ser Ser Ala Lys His Pro Ser Val Leu Glu Arg
1105                1110                1115                1120
Gln Ile Gly Ala Ile Ser Gln Gly Met Ser Val Gln Leu His Val Pro
                1125                1130                1135
Tyr Ser Glu His Ala Lys Ala Pro Val Gly Pro Val Thr Met Gly Leu
            1140                1145                1150
Pro Leu Pro Met Asp Pro Lys Lys Leu Ala Pro Phe Ser Gly Val Lys
        1155                1160                1165
Gln Glu Gln Leu Ser Pro Arg Gly Gln Ala Gly Pro Pro Glu Ser Leu
    1170                1175                1180
Gly Val Pro Thr Ala Gln Glu Ala Ser Val Leu Arg Gly Thr Ala Leu
1185                1190                1195                1200
Gly Ser Val Pro Gly Gly Ser Ile Thr Lys Gly Ile Pro Ser Thr Arg
                1205                1210                1215
Val Pro Ser Asp Ser Ala Ile Thr Tyr Arg Gly Ser Ile Thr His Gly
            1220                1225                1230
Thr Pro Ala Asp Val Leu Tyr Lys Gly Thr Ile Thr Arg Ile Ile Gly
        1235                1240                1245
Glu Asp Ser Pro Ser Arg Leu Asp Arg Gly Arg Glu Asp Ser Leu Pro
    1250                1255                1260
Lys Gly His Val Ile Tyr Glu Gly Lys Lys Gly His Val Leu Ser Tyr
1265                1270                1275                1280
Glu Gly Gly Met Ser Val Thr Gln Cys Ser Lys Glu Asp Gly Arg Ser
                1285                1290                1295
Ser Ser Gly Pro Pro His Glu Thr Ala Ala Pro Lys Arg Thr Tyr Asp
            1300                1305                1310
Met Met Glu Gly Arg Val Gly Arg Ala Ile Ser Ser Ala Ser Ile Glu
        1315                1320                1325
Gly Leu Met Gly Arg Ala Ile Pro Pro Glu Arg His Ser Pro His His
    1330                1335                1340
Leu Lys Glu Gln His His Ile Arg Gly Ser Ile Thr Gln Gly Ile Pro
1345                1350                1355                1360
Arg Ser Tyr Val Glu Ala Gln Glu Asp Tyr Leu Arg Arg Glu Ala Lys
                1365                1370                1375
Leu Leu Lys Arg Glu Gly Thr Pro Pro Pro Pro Pro Pro Ser Arg Asp
            1380                1385                1390
Leu Thr Glu Ala Tyr Lys Thr Gln Ala Leu Gly Pro Leu Lys Leu Lys
        1395                1400                1405
```

Pro Ala His Glu Gly Leu Val Ala Thr Val Lys Glu Ala Gly Arg Ser
1410 1415 1420

Ile His Glu Ile Pro Arg Glu Glu Leu Arg His Thr Pro Glu Leu Pro
1425 1430 1435 1440

Leu Ala Pro Arg Pro Leu Lys Glu Gly Ser Ile Thr Gln Gly Thr Pro
1445 1450 1455

Leu Lys Tyr Asp Thr Gly Ala Ser Thr Thr Gly Ser Lys Lys His Asp
1460 1465 1470

Val Arg Ser Leu Ile Gly Ser Pro Gly Arg Thr Phe Pro Pro Val His
1475 1480 1485

Pro Leu Asp Val Met Ala Asp Ala Arg Ala Leu Glu Arg Ala Cys Tyr
1490 1495 1500

Glu Glu Ser Leu Lys Ser Arg Pro Gly Thr Ala Ser Ser Ser Gly Gly
1505 1510 1515 1520

Ser Ile Ala Arg Gly Ala Pro Val Ile Val Pro Glu Leu Gly Lys Pro
1525 1530 1535

Arg Gln Ser Pro Leu Thr Tyr Glu Asp His Gly Ala Pro Phe Ala Gly
1540 1545 1550

His Leu Pro Arg Gly Ser Pro Val Thr Met Arg Glu Pro Thr Pro Arg
1555 1560 1565

Leu Gln Glu Gly Ser Leu Ser Ser Ser Lys Ala Ser Gln Asp Arg Lys
1570 1575 1580

Leu Thr Ser Thr Pro Arg Glu Ile Ala Lys Ser Pro His Ser Thr Val
1585 1590 1595 1600

Pro Glu His His Pro His Pro Ile Ser Pro Tyr Glu His Leu Leu Arg
1605 1610 1615

Gly Val Ser Gly Val Asp Leu Tyr Arg Ser His Ile Pro Leu Ala Phe
1620 1625 1630

Asp Pro Thr Ser Ile Pro Arg Gly Ile Pro Leu Asp Ala Ala Ala Ala
1635 1640 1645

Tyr Tyr Leu Pro Arg His Leu Ala Pro Asn Pro Thr Tyr Pro His Leu
1650 1655 1660

Tyr Pro Pro Tyr Leu Ile Arg Gly Tyr Pro Asp Thr Ala Ala Leu Glu
1665 1670 1675 1680

Asn Arg Gln Thr Ile Ile Asn Asp Tyr Ile Thr Ser Gln Gln Met His
1685 1690 1695

His Asn Thr Ala Thr Ala Met Ala Gln Arg Ala Asp Met Leu Arg Gly
1700 1705 1710

Leu Ser Pro Arg Glu Ser Ser Leu Ala Leu Asn Tyr Ala Ala Gly Pro
1715 1720 1725

Arg Gly Ile Ile Asp Leu Ser Gln Val Pro His Leu Pro Val Leu Val
1730 1735 1740

Pro Pro Thr Pro Gly Thr Pro Ala Thr Ala Met Asp Arg Leu Ala Tyr
1745 1750 1755 1760

Leu Pro Thr Ala Pro Gln Pro Phe Ser Ser Arg His Ser Ser Ser Pro
1765 1770 1775

Leu Ser Pro Gly Gly Pro Thr His Leu Thr Lys Pro Thr Thr Thr Ser
1780 1785 1790

Ser Ser Glu Arg Glu Arg Asp Arg Asp Arg Glu Arg Asp Arg Asp Arg
1795 1800 1805

Glu Arg Glu Lys Ser Ile Leu Thr Ser Thr Thr Thr Val Glu His Ala
1810 1815 1820

Pro Ile Trp Arg Pro Gly Thr Glu Gln Ser Ser Gly Ser Ser Gly Ser
1825 1830 1835 1840

Ser Gly Gly Gly Gly Gly Ser Ser Ser Arg Pro Ala Ser His Ser His
1845 1850 1855

Ala His Gln His Ser Pro Ile Ser Pro Arg Thr Gln Asp Ala Leu Gln
1860 1865 1870

Gln Arg Pro Ser Val Leu His Asn Thr Gly Met Lys Gly Ile Ile Thr
1875 1880 1885

Ala Val Glu Pro Ser Lys Pro Thr Val Leu Arg Ser Thr Ser Thr Ser
1890 1895 1900
Ser Pro Val Arg Pro Ala Ala Thr Phe Pro Pro Ala Thr His Cys Pro
1905 1910 1915 1920
Leu Gly Gly Thr Leu Asp Gly Val Tyr Pro Thr Leu Met Glu Pro Val
1925 1930 1935
Leu Leu Pro Lys Glu Ala Pro Arg Val Ala Arg Pro Glu Arg Pro Arg
1940 1945 1950
Ala Asp Thr Gly His Ala Phe Leu Ala Lys Pro Pro Ala Arg Ser Gly
1955 1960 1965
Leu Glu Pro Ala Ser Ser Pro Ser Lys Gly Ser Glu Pro Arg Pro Leu
1970 1975 1980
Val Pro Pro Val Ser Gly His Ala Thr Ile Ala Arg Thr Pro Ala Lys
1985 1990 1995 2000
Asn Leu Ala Pro His His Ala Ser Pro Asp Pro Pro Ala Pro Pro Ala
2005 2010 2015
Ser Ala Ser Asp Pro His Arg Glu Lys Thr Gln Ser Lys Pro Phe Ser
2020 2025 2030
Ile Gln Glu Leu Glu Leu Arg Ser Leu Gly Tyr His Gly Ser Ser Tyr
2035 2040 2045
Ser Pro Glu Gly Val Glu Pro Val Ser Pro Val Ser Ser Pro Ser Leu
2050 2055 2060
Thr His Asp Lys Gly Leu Pro Lys His Leu Glu Glu Leu Asp Lys Ser
2065 2070 2075 2080
His Leu Glu Gly Glu Leu Arg Pro Lys Gln Pro Gly Pro Val Lys Leu
2085 2090 2095
Gly Gly Glu Ala Ala His Leu Pro His Leu Arg Pro Leu Pro Glu Ser
2100 2105 2110
Gln Pro Ser Ser Ser Pro Leu Leu Gln Thr Ala Pro Gly Val Lys Gly
2115 2120 2125
His Gln Arg Val Val Thr Leu Ala Gln His Ile Ser Glu Val Ile Thr
2130 2135 2140
Gln Asp Tyr Thr Arg His His Pro Gln Gln Leu Ser Ala Pro Leu Pro
2145 2150 2155 2160
Ala Pro Leu Tyr Ser Phe Pro Gly Ala Ser Cys Pro Val Leu Asp Leu
2165 2170 2175
Arg Arg Pro Pro Ser Asp Leu Tyr Leu Pro Pro Pro Asp His Gly Ala
2180 2185 2190
Pro Ala Arg Gly Ser Pro His Ser Glu Gly Gly Lys Arg Ser Pro Glu
2195 2200 2205
Pro Asn Lys Thr Ser Val Leu Gly Gly Gly Glu Asp Gly Ile Glu Pro
2210 2215 2220
Val Ser Pro Pro Glu Gly Met Thr Glu Pro Gly His Ser Arg Ser Ala
2225 2230 2235 2240
Val Tyr Pro Leu Leu Tyr Arg Asp Gly Glu Gln Thr Glu Pro Ser Arg
2245 2250 2255
Met Gly Ser Lys Ser Pro Gly Asn Thr Ser Gln Pro Pro Ala Phe Phe
2260 2265 2270
Ser Lys Leu Thr Glu Ser Asn Ser Ala Met Val Lys Ser Lys Lys Gln
2275 2280 2285
Glu Ile Asn Lys Lys Leu Asn Thr His Asn Arg Asn Glu Pro Glu Tyr
2290 2295 2300
Asn Ile Ser Gln Pro Gly Thr Glu Ile Phe Asn Met Pro Ala Ile Thr
2305 2310 2315 2320
Gly Thr Gly Leu Met Thr Tyr Arg Ser Gln Ala Val Gln Glu His Ala
2325 2330 2335
Ser Thr Asn Met Gly Leu Glu Ala Ile Ile Arg Lys Ala Leu Met Gly
2340 2345 2350
Lys Tyr Asp Gln Trp Glu Glu Ser Pro Pro Leu Ser Ala Asn Ala Phe
2355 2360 2365

```
Asn Pro Leu Asn Ala Ser Ala Ser Leu Pro Ala Ala Met Pro Ile Thr
    2370                2375                2380
Ala Ala Asp Gly Arg Ser Asp His Thr Leu Thr Ser Pro Gly Gly Gly
2385                2390                2395                2400
Gly Lys Ala Lys Val Ser Gly Arg Pro Ser Ser Arg Lys Ala Lys Ser
                2405                2410                2415
Pro Ala Pro Gly Leu Ala Ser Gly Asp Arg Pro Pro Ser Val Ser Ser
            2420                2425                2430
Val His Ser Glu Gly Asp Cys Asn Arg Arg Thr Pro Leu Thr Asn Arg
        2435                2440                2445
Val Trp Glu Asp Arg Pro Ser Ser Ala Gly Ser Thr Pro Phe Pro Tyr
    2450                2455                2460
Asn Pro Leu Ile Met Arg Leu Gln Ala Gly Val Met Ala Ser Pro Pro
2465                2470                2475                2480
Pro Pro Gly Leu Pro Ala Gly Ser Gly Pro Leu Ala Gly Pro His His
                2485                2490                2495
Ala Trp Asp Glu Glu Pro Lys Pro Leu Leu Cys Ser Gln Tyr Glu Thr
            2500                2505                2510
Leu Ser Asp Ser Glu
        2515

<210> 6
<211> 8388
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (626)...(8047)

<221> misc_feature
<222> (1)...(8388)
<223> n = A,T,C or G

<400> 6
cttaaaaaaa aaacccttac ttgtggttaa aggaaaagaa ataaagactt aggaaaaatg      60
taattttcca gggggtacct acacccaaga catatggttc tcaagaggna ctcagcatat     120
cactttgatt ccagagaagc tacaaaagtc attaccaaac tccaggctgg aaagcagtgc     180
tcatactaaa tatttaaaca tttaaagacc tgattaagag acatcaaagg ctttatacca     240
ggggcacacc aacagagaca ggctttttca aggataattt atgtctgccc attgtcttct     300
ggcttaggag acatagaggg aaacatcacc taggaaaacc agtaaccaat gtgtaccatc     360
caggagttat tctatgacaa aaccaaaagt tttgttcttg tgtacttctc tgtgcaccat     420
ctttctatat ctatttagaa aacaaaacaa atttggtaa cacgcttgtg tataaagagc      480
aggacagcgg tgtcacagat caacctagaa agtaattatt taacgagtaa atgactcata     540
taggacaagg caagctgtga ctttcaacct gttctgtctc gtgccgaatt cggcacgagc     600
caaagcctac ctggacccta ccacc atg tca gga tcc aca cag cct gtg gca       652
                            Met Ser Gly Ser Thr Gln Pro Val Ala
                              1               5

cag aca tgg cgg gct gct gag ccc cgc tac cca ccc cat ggc atc tcc       700
Gln Thr Trp Arg Ala Ala Glu Pro Arg Tyr Pro Pro His Gly Ile Ser
 10                  15                  20                  25

tac ccg gtg cag ata gcc cgg tcc cac acg gac gtg ggg ctg ctt gag       748
Tyr Pro Val Gln Ile Ala Arg Ser His Thr Asp Val Gly Leu Leu Glu
                 30                  35                  40

tac caa cac cac ccc cgt gac tac acc tca cac ctg tca ccc ggt tcc       796
Tyr Gln His His Pro Arg Asp Tyr Thr Ser His Leu Ser Pro Gly Ser
             45                  50                  55
```

atc atc cag cca cag agg agg cgg ccc tca ctg ctg tca gag ttc cag 844
Ile Ile Gln Pro Gln Arg Arg Arg Pro Ser Leu Leu Ser Glu Phe Gln
60 65 70

cct ggg agt gaa cgg tct cag gag ctc cac ctg cgc cct gag tcc cgc 892
Pro Gly Ser Glu Arg Ser Gln Glu Leu His Leu Arg Pro Glu Ser Arg
75 80 85

acg ttc ctg cct gag ctg ggc aag ccc gac ata gaa ttc acc gag agc 940
Thr Phe Leu Pro Glu Leu Gly Lys Pro Asp Ile Glu Phe Thr Glu Ser
90 95 100 105

aag cgc ccc cgc ctg gag cta cta ccc gat acc ctg ctg cgc cca tca 988
Lys Arg Pro Arg Leu Glu Leu Leu Pro Asp Thr Leu Leu Arg Pro Ser
110 115 120

ccc ctg ctg gcc act ggg cag ccg agt ggg tct gaa gac ctt acc aag 1036
Pro Leu Leu Ala Thr Gly Gln Pro Ser Gly Ser Glu Asp Leu Thr Lys
125 130 135

gac cgt agc ctg gca ggc aag ctg gag cct gtg tca cct ccc agt ccc 1084
Asp Arg Ser Leu Ala Gly Lys Leu Glu Pro Val Ser Pro Pro Ser Pro
140 145 150

ccg cac gct gac cct gag cta gag ctg gcg cca tct cga ctg tcc aag 1132
Pro His Ala Asp Pro Glu Leu Glu Leu Ala Pro Ser Arg Leu Ser Lys
155 160 165

gag gag ctg atc cag aac atg gac cgc gtg gac cgt gag atc acc atg 1180
Glu Glu Leu Ile Gln Asn Met Asp Arg Val Asp Arg Glu Ile Thr Met
170 175 180 185

gta gag cag cag atc tcc aag ctg aag aag aag cag caa cag ttg gag 1228
Val Glu Gln Gln Ile Ser Lys Leu Lys Lys Lys Gln Gln Gln Leu Glu
190 195 200

gag gag gcc gcc aag ccg ccc gaa ccc gag aag cct gtg tcg cca cca 1276
Glu Glu Ala Ala Lys Pro Pro Glu Pro Glu Lys Pro Val Ser Pro Pro
205 210 215

ccc ata gaa tca aag cac cga agc ctg gtc cag atc atc tac gat gag 1324
Pro Ile Glu Ser Lys His Arg Ser Leu Val Gln Ile Ile Tyr Asp Glu
220 225 230

aac cgg aag aaa gcc gaa gcc gca cac cgg atc cta gaa ggc ctg ggg 1372
Asn Arg Lys Lys Ala Glu Ala Ala His Arg Ile Leu Glu Gly Leu Gly
235 240 245

ccc cag gtg gag ctg cct ctg tac aac cag ccg tct gac aca cgc cag 1420
Pro Gln Val Glu Leu Pro Leu Tyr Asn Gln Pro Ser Asp Thr Arg Gln
250 255 260 265

tac cat gaa aac atc aaa ata aac cag gcg atg cgg aag aag ctg atc 1468
Tyr His Glu Asn Ile Lys Ile Asn Gln Ala Met Arg Lys Lys Leu Ile
270 275 280

ttg tac ttt aag cgg agg aac cac gcg cgc aag cag tgg gaa cag cgc 1516
Leu Tyr Phe Lys Arg Arg Asn His Ala Arg Lys Gln Trp Glu Gln Arg
285 290 295

```
ttc tgc cag cgc tat gac cag ctc atg gag gcg tgg gag aag aag gta      1564
Phe Cys Gln Arg Tyr Asp Gln Leu Met Glu Ala Trp Glu Lys Lys Val
        300                 305                 310

gag cgc ata gag aac aat ccg cga agg agg gcc aag gag agc aag gtg      1612
Glu Arg Ile Glu Asn Asn Pro Arg Arg Arg Ala Lys Glu Ser Lys Val
    315                 320                 325

agg gag tac tac gag aaa cag ttc ccg gag atc cgc aag cag cgg gag      1660
Arg Glu Tyr Tyr Glu Lys Gln Phe Pro Glu Ile Arg Lys Gln Arg Glu
330                 335                 340                 345

ctg cag gag cgc atg cag agc agg gtg ggc cag cgt ggc agt ggg ctc      1708
Leu Gln Glu Arg Met Gln Ser Arg Val Gly Gln Arg Gly Ser Gly Leu
                350                 355                 360

tcc atg tcg gct gcc cgc agt gag cat gag gtt tct gag atc att gat      1756
Ser Met Ser Ala Ala Arg Ser Glu His Glu Val Ser Glu Ile Ile Asp
            365                 370                 375

ggc ttg tct gag cag gag aac ctg gag aag cag atg cgc cag ctg gcc      1804
Gly Leu Ser Glu Gln Glu Asn Leu Glu Lys Gln Met Arg Gln Leu Ala
        380                 385                 390

gtg atc ccg ccc atg ttg tac gac gcg gac cag cag agg atc aag ttc      1852
Val Ile Pro Pro Met Leu Tyr Asp Ala Asp Gln Gln Arg Ile Lys Phe
    395                 400                 405

atc aac atg aat gga ctc atg gat gac ccc atg aag gtc tac aag gac      1900
Ile Asn Met Asn Gly Leu Met Asp Asp Pro Met Lys Val Tyr Lys Asp
410                 415                 420                 425

cgt cag gtt acc aac atg tgg agc gag cag gag agg gac acc ttc cgt      1948
Arg Gln Val Thr Asn Met Trp Ser Glu Gln Glu Arg Asp Thr Phe Arg
                430                 435                 440

gag aag ttt atg cag cac cct aag aac ttt ggc ctg att gcc tca ttc      1996
Glu Lys Phe Met Gln His Pro Lys Asn Phe Gly Leu Ile Ala Ser Phe
            445                 450                 455

ctg gag aga aag acg gtc gct gag tgt gtc ctc tat tac tac ctg acc      2044
Leu Glu Arg Lys Thr Val Ala Glu Cys Val Leu Tyr Tyr Tyr Leu Thr
        460                 465                 470

aag aag aat gaa aat tac aag agc ttg gtg agg cgg agc tat cgg cgc      2092
Lys Lys Asn Glu Asn Tyr Lys Ser Leu Val Arg Arg Ser Tyr Arg Arg
    475                 480                 485

cgt ggc aag agc cag cag cag cag cag cag caa caa cag cag cag cag      2140
Arg Gly Lys Ser Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
490                 495                 500                 505

cag cag atg gca cgg agc agc cag gag gag aag gag gag aag gag aag      2188
Gln Gln Met Ala Arg Ser Ser Gln Glu Glu Lys Glu Glu Lys Glu Lys
                510                 515                 520

gag aag gag gcc gac aag gag gaa gag aag cag gat gcg gag aac gag      2236
Glu Lys Glu Ala Asp Lys Glu Glu Glu Lys Gln Asp Ala Glu Asn Glu
            525                 530                 535
```

aag gaa gaa ctc agc aag gag aag aca gac gac act tct ggc gag gac 2284
Lys Glu Glu Leu Ser Lys Glu Lys Thr Asp Asp Thr Ser Gly Glu Asp
540 545 550

aac cat gag aaa gag gcc gtg gcc tcc aaa ggc cgc aaa act gcc aac 2332
Asn His Glu Lys Glu Ala Val Ala Ser Lys Gly Arg Lys Thr Ala Asn
555 560 565

agc caa ggc cgc cgc aaa ggc cgt atc acg cgc tcc atg gcc aac gag 2380
Ser Gln Gly Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Ala Asn Glu
570 575 580 585

gcc aac cat gag gag aca gcc acc cca cag caa agt tca gag ctg gct 2428
Ala Asn His Glu Glu Thr Ala Thr Pro Gln Gln Ser Ser Glu Leu Ala
590 595 600

tcc atg gag atg aac gag agt tct cgc tgg act gag gaa gag atg gag 2476
Ser Met Glu Met Asn Glu Ser Ser Arg Trp Thr Glu Glu Glu Met Glu
605 610 615

aca gca aag aaa ggc ctc ctg gaa cat ggg agg aac tgg tca gcc att 2524
Thr Ala Lys Lys Gly Leu Leu Glu His Gly Arg Asn Trp Ser Ala Ile
620 625 630

gcc cgc atg gtg ggc tcc aag acc gtg tcc cag tgt aag aac ttc tac 2572
Ala Arg Met Val Gly Ser Lys Thr Val Ser Gln Cys Lys Asn Phe Tyr
635 640 645

ttc aac tac aag aag agg cag aac ctg gac gaa atc ctt cag cag cac 2620
Phe Asn Tyr Lys Lys Arg Gln Asn Leu Asp Glu Ile Leu Gln Gln His
650 655 660 665

aag cta aag atg gag aag gag agg aac gct cgg agg aag aag aag aag 2668
Lys Leu Lys Met Glu Lys Glu Arg Asn Ala Arg Arg Lys Lys Lys Lys
670 675 680

acc cca gct gcg gcg agc gag gag aca gcc ttc cca cct gcc gct gag 2716
Thr Pro Ala Ala Ala Ser Glu Glu Thr Ala Phe Pro Pro Ala Ala Glu
685 690 695

gac gaa gag atg gaa gca tca ggc gca agt gcc aat gag gaa gag ctg 2764
Asp Glu Glu Met Glu Ala Ser Gly Ala Ser Ala Asn Glu Glu Glu Leu
700 705 710

gcg gag gag gca gaa gcc tca cag gcc tct ggg aat gag gtt ccc aga 2812
Ala Glu Glu Ala Glu Ala Ser Gln Ala Ser Gly Asn Glu Val Pro Arg
715 720 725

gtt ggg gag tgc agt ggc cca gct gct gtc aac aac agc tct gat act 2860
Val Gly Glu Cys Ser Gly Pro Ala Ala Val Asn Asn Ser Ser Asp Thr
730 735 740 745

gag agt gtc cca tcc ccg cgt tca gaa gcc atg aag gac act ggg cct 2908
Glu Ser Val Pro Ser Pro Arg Ser Glu Ala Met Lys Asp Thr Gly Pro
750 755 760

aaa ccc act ggc act gaa gca ttg ccc gct gcc acc cag cca cct gtt 2956
Lys Pro Thr Gly Thr Glu Ala Leu Pro Ala Ala Thr Gln Pro Pro Val
765 770 775

```
cct cct cca gaa gaa ccg gca gta gcc cct gct gag ccc tcc cca gtc      3004
Pro Pro Pro Glu Glu Pro Ala Val Ala Pro Ala Glu Pro Ser Pro Val
            780                 785                 790

cct gat gcc agt ggc cca cca tcc cca gag cct tcc cat cac ctg ccg      3052
Pro Asp Ala Ser Gly Pro Pro Ser Pro Glu Pro Ser His His Leu Pro
    795                 800                 805

cac ccc cgg cta ctg tgg aca agg atg aac aag aag ccc cgg ctg ctc      3100
His Pro Arg Leu Leu Trp Thr Arg Met Asn Lys Lys Pro Arg Leu Leu
810                 815                 820                 825

cag ctc ccc aga cag agg atg cca agg agc aga agt ctg agg ccg agg      3148
Gln Leu Pro Arg Gln Arg Met Pro Arg Ser Arg Ser Leu Arg Pro Arg
                830                 835                 840

aga tcg atg tgg gaa aag cca gag gag ccc gag gcc tct gag gag ccc      3196
Arg Ser Met Trp Glu Lys Pro Glu Glu Pro Glu Ala Ser Glu Glu Pro
            845                 850                 855

ccg gag agt gta aag agt gac cac aag gag gag acc gag gaa gag cct      3244
Pro Glu Ser Val Lys Ser Asp His Lys Glu Glu Thr Glu Glu Glu Pro
        860                 865                 870

gaa gac aaa gcc aag ggc aca gag gcc att gaa act gtg tct gag gca      3292
Glu Asp Lys Ala Lys Gly Thr Glu Ala Ile Glu Thr Val Ser Glu Ala
    875                 880                 885

cca ctt aag gtg gag gag gct ggt agc aag gca gct gtg acc aag ggt      3340
Pro Leu Lys Val Glu Glu Ala Gly Ser Lys Ala Ala Val Thr Lys Gly
890                 895                 900                 905

tcc agc tca ggt gcc acc cag gac agt gac ttc agt gcc acc tgc agt      3388
Ser Ser Ser Gly Ala Thr Gln Asp Ser Asp Phe Ser Ala Thr Cys Ser
                910                 915                 920

gcc gat gag gtg gac gaa ccc gaa gga ggt gac aag ggc agg ctg ctg      3436
Ala Asp Glu Val Asp Glu Pro Glu Gly Gly Asp Lys Gly Arg Leu Leu
            925                 930                 935

tca cca agg ccc agc ctc ctc acc ccg gct gga gat ccc cgg gcc agt      3484
Ser Pro Arg Pro Ser Leu Leu Thr Pro Ala Gly Asp Pro Arg Ala Ser
        940                 945                 950

acc tcg ccc cag aag ccg ctg gac ctg aag cag ctg aag cag cga gca      3532
Thr Ser Pro Gln Lys Pro Leu Asp Leu Lys Gln Leu Lys Gln Arg Ala
    955                 960                 965

gcc gcc atc ccc cct atc cag gtc acc aag gtc cat gag ccc ccc cgg      3580
Ala Ala Ile Pro Pro Ile Gln Val Thr Lys Val His Glu Pro Pro Arg
970                 975                 980                 985

gag gac aca gta ccc cca aag cca gtt ccc cct gtg cct cca ccc acg      3628
Glu Asp Thr Val Pro Pro Lys Pro Val Pro Pro Val Pro Pro Pro Thr
                990                 995                 1000

cag cac cta cag cca gag ggt gac gtg tct cag cag tcg gga gga agt      3676
Gln His Leu Gln Pro Glu Gly Asp Val Ser Gln Gln Ser Gly Gly Ser
            1005                1010                1015
```

```
cca cgt ggc aag tcc cgc agc cca gtg cct cct gcc gag aaa gag gca      3724
Pro Arg Gly Lys Ser Arg Ser Pro Val Pro Pro Ala Glu Lys Glu Ala
        1020                1025                1030

gag aaa ccc gca ttc ttt ccg gct ttc cca act gag ggc cca aag cta      3772
Glu Lys Pro Ala Phe Phe Pro Ala Phe Pro Thr Glu Gly Pro Lys Leu
    1035                1040                1045

ccg act gag ccc cca cgc tgg tca tcg ggc ctg ccc ttc ccc atc cct      3820
Pro Thr Glu Pro Pro Arg Trp Ser Ser Gly Leu Pro Phe Pro Ile Pro
1050                1055                1060                1065

cca cgg gag gtg atc aag act tcc cca cac gcc gct gac ccc tct gcc      3868
Pro Arg Glu Val Ile Lys Thr Ser Pro His Ala Ala Asp Pro Ser Ala
                1070                1075                1080

ttc tcc tac aca ccc ccc ggt cac ccg ctg cct ctg ggc ctc cac gat      3916
Phe Ser Tyr Thr Pro Pro Gly His Pro Leu Pro Leu Gly Leu His Asp
            1085                1090                1095

agt gcc cgg ccc gtc ctg cca cgt ccc ccc atc tct aac ccc cca ccc      3964
Ser Ala Arg Pro Val Leu Pro Arg Pro Pro Ile Ser Asn Pro Pro Pro
        1100                1105                1110

ctc atc tcc tct gcc aag cat ccc ggc gta ctt gag agg cag ctg ggt      4012
Leu Ile Ser Ser Ala Lys His Pro Gly Val Leu Glu Arg Gln Leu Gly
    1115                1120                1125

gcc atc tcc cag cag ggg atg tca gtc cag ctt cgt gtg cct cac tca      4060
Ala Ile Ser Gln Gln Gly Met Ser Val Gln Leu Arg Val Pro His Ser
1130                1135                1140                1145

gag cat gcc aag gcc ccc atg ggc cct ctc acc atg ggg ctg ccc ctt      4108
Glu His Ala Lys Ala Pro Met Gly Pro Leu Thr Met Gly Leu Pro Leu
                1150                1155                1160

gcc gtg gac cct aag aag ctg ggg aca gca ctg ggc tcc gcc acc agt      4156
Ala Val Asp Pro Lys Lys Leu Gly Thr Ala Leu Gly Ser Ala Thr Ser
            1165                1170                1175

gga agc atc acc aag ggc ctc ccc agt acc cgg gct gca gac ggc ccc      4204
Gly Ser Ile Thr Lys Gly Leu Pro Ser Thr Arg Ala Ala Asp Gly Pro
        1180                1185                1190

agc tac aga ggc tct atc acc cac ggc acg ccc gca gac gtc ctc tac      4252
Ser Tyr Arg Gly Ser Ile Thr His Gly Thr Pro Ala Asp Val Leu Tyr
    1195                1200                1205

aag ggt acc atc agc agg atc gtc ggt gag gac agc cca agt cgc ctt      4300
Lys Gly Thr Ile Ser Arg Ile Val Gly Glu Asp Ser Pro Ser Arg Leu
1210                1215                1220                1225

gac cgg gca cga gag gac acc ctg ccc aag ggc cat gtc atc tat gag      4348
Asp Arg Ala Arg Glu Asp Thr Leu Pro Lys Gly His Val Ile Tyr Glu
                1230                1235                1240

ggc aag aaa ggc cac gtc cta tcc tat gaa ggt ggt atg tcc gtg tca      4396
Gly Lys Lys Gly His Val Leu Ser Tyr Glu Gly Gly Met Ser Val Ser
            1245                1250                1255
```

cag tgc tct aag gag gat gga agg agc agc tcg ggc cca ccc cat gag 4444
Gln Cys Ser Lys Glu Asp Gly Arg Ser Ser Ser Gly Pro Pro His Glu
1260 1265 1270

act gcc gcc cct aaa cgc acc tat gac atg atg gag ggc cgt gta ggc 4492
Thr Ala Ala Pro Lys Arg Thr Tyr Asp Met Met Glu Gly Arg Val Gly
1275 1280 1285

agg act gtc acc tca gcc agc ata gag gga ctc atg ggc cgc gcc atc 4540
Arg Thr Val Thr Ser Ala Ser Ile Glu Gly Leu Met Gly Arg Ala Ile
1290 1295 1300 1305

cct gag cag cac agc ccc cac ctc aag gag cag cat cac atc cga ggc 4588
Pro Glu Gln His Ser Pro His Leu Lys Glu Gln His His Ile Arg Gly
1310 1315 1320

tcc atc acg caa ggc atc ccg agg tcc tat gtg gag gcg cag gag gac 4636
Ser Ile Thr Gln Gly Ile Pro Arg Ser Tyr Val Glu Ala Gln Glu Asp
1325 1330 1335

tac tta cgg cgg gag gcc aag ctc ttg aag cga gaa ggg aca cca cca 4684
Tyr Leu Arg Arg Glu Ala Lys Leu Leu Lys Arg Glu Gly Thr Pro Pro
1340 1345 1350

ccc cca cca cca cct cgg gac ctg act gag acc tac aag ccc cgg ccc 4732
Pro Pro Pro Pro Pro Arg Asp Leu Thr Glu Thr Tyr Lys Pro Arg Pro
1355 1360 1365

ctg gac cct ctg ggt ccc ctg aag ctg aag ccg act cac gag ggt gtg 4780
Leu Asp Pro Leu Gly Pro Leu Lys Leu Lys Pro Thr His Glu Gly Val
1370 1375 1380 1385

gta gca act gtg aag gag gcg ggc cgc tct atc cat gag atc ccg aga 4828
Val Ala Thr Val Lys Glu Ala Gly Arg Ser Ile His Glu Ile Pro Arg
1390 1395 1400

gag gag ctg cgc cgc aca cct gag cta ccc ctg gca cca cgg cct ctg 4876
Glu Glu Leu Arg Arg Thr Pro Glu Leu Pro Leu Ala Pro Arg Pro Leu
1405 1410 1415

aag gag ggt tcc atc acc cag ggc acc cca ctc aag tac gac tct ggg 4924
Lys Glu Gly Ser Ile Thr Gln Gly Thr Pro Leu Lys Tyr Asp Ser Gly
1420 1425 1430

gca ccc tcc act ggc acc aag aaa cac gac gtg cgc tcc atc atc ggc 4972
Ala Pro Ser Thr Gly Thr Lys Lys His Asp Val Arg Ser Ile Ile Gly
1435 1440 1445

agc ccc ggc cgg cct ttc cct gcc ctg cac ccg ctg gac ata atg gct 5020
Ser Pro Gly Arg Pro Phe Pro Ala Leu His Pro Leu Asp Ile Met Ala
1450 1455 1460 1465

gac gcc cgg gca ctg gag cgt gcc tgc tat gaa gag agt ctg aag agc 5068
Asp Ala Arg Ala Leu Glu Arg Ala Cys Tyr Glu Glu Ser Leu Lys Ser
1470 1475 1480

cgg tca ggg acc agc agt ggt gca ggg ggc tcc atc aca cgt ggg gct 5116
Arg Ser Gly Thr Ser Ser Gly Ala Gly Gly Ser Ile Thr Arg Gly Ala
1485 1490 1495

```
cca gtc gtc gtg cct gaa ctg ggc aag cca cgg caa agc cca ctg act      5164
Pro Val Val Val Pro Glu Leu Gly Lys Pro Arg Gln Ser Pro Leu Thr
        1500                1505                1510

tac gaa gac cac ggg gca ccc ttc acc agt cac ctg cca cgt ggc tcc      5212
Tyr Glu Asp His Gly Ala Pro Phe Thr Ser His Leu Pro Arg Gly Ser
    1515                1520                1525

cct gtg acc acg agg gag ccc acg cca cgc ctt cag gaa ggc agc ctc      5260
Pro Val Thr Thr Arg Glu Pro Thr Pro Arg Leu Gln Glu Gly Ser Leu
1530                1535                1540                1545

cta tcc agc aag gcg tcc cag gac cgg aag ctg aca tct aca ccc cgg      5308
Leu Ser Ser Lys Ala Ser Gln Asp Arg Lys Leu Thr Ser Thr Pro Arg
                1550                1555                1560

gag atc gcc aag tcc cca cac agc act gtg ccc gag cac cac cct cac      5356
Glu Ile Ala Lys Ser Pro His Ser Thr Val Pro Glu His His Pro His
            1565                1570                1575

ccc atc tcc ccc tat gag cac ttg ctc cgg ggc gtg act ggt gtg gac      5404
Pro Ile Ser Pro Tyr Glu His Leu Leu Arg Gly Val Thr Gly Val Asp
        1580                1585                1590

ctg tac cgt ggt cac atc cca ttg gcc ttt gac ccc acc tcc ata ccc      5452
Leu Tyr Arg Gly His Ile Pro Leu Ala Phe Asp Pro Thr Ser Ile Pro
    1595                1600                1605

cga ggg atc cct ctg gaa gca gca gcc gca gcc tac tac ctg ccc cgg      5500
Arg Gly Ile Pro Leu Glu Ala Ala Ala Ala Ala Tyr Tyr Leu Pro Arg
1610                1615                1620                1625

cac ttg gcc ccc agc ccc acc tac cca cac ctg tac cca cct tac ctc      5548
His Leu Ala Pro Ser Pro Thr Tyr Pro His Leu Tyr Pro Pro Tyr Leu
                1630                1635                1640

atc cgc ggc tac cct gac acg gcg gcc ctg gag aac cgc cag acc atc      5596
Ile Arg Gly Tyr Pro Asp Thr Ala Ala Leu Glu Asn Arg Gln Thr Ile
            1645                1650                1655

atc aat gac tac atc acc tcg cag cag atg cac cac aac gct gcc tcc      5644
Ile Asn Asp Tyr Ile Thr Ser Gln Gln Met His His Asn Ala Ala Ser
        1660                1665                1670

gcc atg gcc cag cgt gct gac atg ctg agg ggt ctg tca ccg cga gag      5692
Ala Met Ala Gln Arg Ala Asp Met Leu Arg Gly Leu Ser Pro Arg Glu
    1675                1680                1685

tcc tcg ctg gcc ctc aat tat gcc gct ggc cca aga ggc att atc gac      5740
Ser Ser Leu Ala Leu Asn Tyr Ala Ala Gly Pro Arg Gly Ile Ile Asp
1690                1695                1700                1705

ctg tcc caa gtg cca cac ctg ccc gtg ctg gtg cca cca acg cca ggc      5788
Leu Ser Gln Val Pro His Leu Pro Val Leu Val Pro Pro Thr Pro Gly
                1710                1715                1720

acc cct gcc acc gcc atc gac cgc ctt gcc tac ctc ccc act gcg ccc      5836
Thr Pro Ala Thr Ala Ile Asp Arg Leu Ala Tyr Leu Pro Thr Ala Pro
            1725                1730                1735
```

cca ccc ttc agc agc cgc cac agt agc tca ccg ctg tcc cca gga ggc 5884
Pro Pro Phe Ser Ser Arg His Ser Ser Ser Pro Leu Ser Pro Gly Gly
1740 1745 1750

ccc act cac cta gct aaa cca act gcc aca tct tca tcg gag cgg gaa 5932
Pro Thr His Leu Ala Lys Pro Thr Ala Thr Ser Ser Ser Glu Arg Glu
1755 1760 1765

cgg gaa cgt gag cgg gaa cga gac aag tcc atc ctc acg tct acc act 5980
Arg Glu Arg Glu Arg Glu Arg Asp Lys Ser Ile Leu Thr Ser Thr Thr
1770 1775 1780 1785

aca gtg gag cat gca ccc atc tgg aga cct ggt acg gag cag agc agc 6028
Thr Val Glu His Ala Pro Ile Trp Arg Pro Gly Thr Glu Gln Ser Ser
1790 1795 1800

ggg gct ggg ggc agc agc cgc ccc gcc tcc cac acc cac cag cac tcg 6076
Gly Ala Gly Gly Ser Ser Arg Pro Ala Ser His Thr His Gln His Ser
1805 1810 1815

ccc atc tcc ccc cgg acc cag gac gcc ttg cag cag agg ccc agt gtg 6124
Pro Ile Ser Pro Arg Thr Gln Asp Ala Leu Gln Gln Arg Pro Ser Val
1820 1825 1830

ctg cac aac acg agc atg aag ggc gtg gtc acc tcc gtg gaa ccc ggc 6172
Leu His Asn Thr Ser Met Lys Gly Val Val Thr Ser Val Glu Pro Gly
1835 1840 1845

acg ccc acg gtc ctg agg tgg gcc agg tcc acc tcc acc tct tcg cct 6220
Thr Pro Thr Val Leu Arg Trp Ala Arg Ser Thr Ser Thr Ser Ser Pro
1850 1855 1860 1865

gtc cgc cca gct gcc aca ttc cca cct gcc acc cac tgc cca ctt ggt 6268
Val Arg Pro Ala Ala Thr Phe Pro Pro Ala Thr His Cys Pro Leu Gly
1870 1875 1880

ggc acc ctt gaa ggg gtc tac cct acc ctc atg gag ccc gtc ctg tta 6316
Gly Thr Leu Glu Gly Val Tyr Pro Thr Leu Met Glu Pro Val Leu Leu
1885 1890 1895

ccc aag gag acc tct cgg gtc gcc cgg ccc gag cgg gcc cgg gtg gac 6364
Pro Lys Glu Thr Ser Arg Val Ala Arg Pro Glu Arg Ala Arg Val Asp
1900 1905 1910

gct ggc cat gcc ttt ctt acc aaa ccc ccg ggc cgg gag ccc gcc tcc 6412
Ala Gly His Ala Phe Leu Thr Lys Pro Pro Gly Arg Glu Pro Ala Ser
1915 1920 1925

tca ccc agc aag agc tcc gag ccc cga tcc cta gca ccc ccc agc tcc 6460
Ser Pro Ser Lys Ser Ser Glu Pro Arg Ser Leu Ala Pro Pro Ser Ser
1930 1935 1940 1945

agc cac aca gcc atc gcc cgc acc cca gca aag aac ctt gca ccc cac 6508
Ser His Thr Ala Ile Ala Arg Thr Pro Ala Lys Asn Leu Ala Pro His
1950 1955 1960

cat gcc agt ccg gac ccg ccg gcg ccc acc tcg gcc tca gat ctg cac 6556
His Ala Ser Pro Asp Pro Pro Ala Pro Thr Ser Ala Ser Asp Leu His
1965 1970 1975

cga gaa aag act caa agt aaa ccc ttt tcc atc cag gaa ttg gaa ctc 6604
Arg Glu Lys Thr Gln Ser Lys Pro Phe Ser Ile Gln Glu Leu Glu Leu
1980 1985 1990

cgt tct ctg ggt tac cac agt gga gct ggc tac agc ccc gat ggg gtg 6652
Arg Ser Leu Gly Tyr His Ser Gly Ala Gly Tyr Ser Pro Asp Gly Val
1995 2000 2005

gag ccc atc agc ccg gtg agc tcc ccc agc ctg acc cac gac aag ggg 6700
Glu Pro Ile Ser Pro Val Ser Ser Pro Ser Leu Thr His Asp Lys Gly
2010 2015 2020 2025

ctc tcc aaa cct ctg gaa gag cta gag aag agc cac ttg gaa ggg gag 6748
Leu Ser Lys Pro Leu Glu Glu Leu Glu Lys Ser His Leu Glu Gly Glu
2030 2035 2040

ctg cgg cac aag cag cca ggc ccc atg aag ctc agc gcg gag gct gcc 6796
Leu Arg His Lys Gln Pro Gly Pro Met Lys Leu Ser Ala Glu Ala Ala
2045 2050 2055

cat ctc cca cat ctg cgg cca ctg ccc gag agc cag ccc tca tcc agc 6844
His Leu Pro His Leu Arg Pro Leu Pro Glu Ser Gln Pro Ser Ser Ser
2060 2065 2070

cca ctc ctc cag act gcc cca ggc atc aaa ggt cac cag agg gtg gtc 6892
Pro Leu Leu Gln Thr Ala Pro Gly Ile Lys Gly His Gln Arg Val Val
2075 2080 2085

acc ctg gct cag cac atc agc gag gtc att acg cag gac tac acg cgc 6940
Thr Leu Ala Gln His Ile Ser Glu Val Ile Thr Gln Asp Tyr Thr Arg
2090 2095 2100 2105

cac cac ccg cag cag ctc agt ggc ccc ctt ccc gcc cct ctc tac tcc 6988
His His Pro Gln Gln Leu Ser Gly Pro Leu Pro Ala Pro Leu Tyr Ser
2110 2115 2120

ttt ccc gga gcc agc tgc cct gtc ctg gat ctt cgc cgc cca ccc agt 7036
Phe Pro Gly Ala Ser Cys Pro Val Leu Asp Leu Arg Arg Pro Pro Ser
2125 2130 2135

gac ctc tac ctc cca ccc ccc gac cat ggc acc cca gcc cgg gga tcc 7084
Asp Leu Tyr Leu Pro Pro Pro Asp His Gly Thr Pro Ala Arg Gly Ser
2140 2145 2150

ccc cac agt gaa ggg ggc aaa agg tcc cca gaa ccc agc aaa aca tcg 7132
Pro His Ser Glu Gly Gly Lys Arg Ser Pro Glu Pro Ser Lys Thr Ser
2155 2160 2165

gtc ctg ggc agc agc gag gat gcc att gag cct gtg tcc cca cca gag 7180
Val Leu Gly Ser Ser Glu Asp Ala Ile Glu Pro Val Ser Pro Pro Glu
2170 2175 2180 2185

ggc atg act gag cca gga cat gct cgg agc act gcg tac cca ctg ctg 7228
Gly Met Thr Glu Pro Gly His Ala Arg Ser Thr Ala Tyr Pro Leu Leu
2190 2195 2200

tat cga gac ggg gaa cag ggc gag ccc agg atg ggt cta gag tct cca 7276
Tyr Arg Asp Gly Glu Gln Gly Glu Pro Arg Met Gly Leu Glu Ser Pro
2205 2210 2215

ggc aac acc agc cag ccg cca acc ttc ttc agt aag ctg act gag agc 7324
Gly Asn Thr Ser Gln Pro Pro Thr Phe Phe Ser Lys Leu Thr Glu Ser
2220 2225 2230

aac tcc gcc atg gtg aag tcg aag aag cag gag atc aac aag aaa ctc 7372
Asn Ser Ala Met Val Lys Ser Lys Lys Gln Glu Ile Asn Lys Lys Leu
2235 2240 2245

aac acc cac aac cgg aac gag cca gaa tac aat att ggc cag cct ggg 7420
Asn Thr His Asn Arg Asn Glu Pro Glu Tyr Asn Ile Gly Gln Pro Gly
2250 2255 2260 2265

acg gaa atc ttc aac atg ccc gcc atc act gga gca ggc ctt atg acc 7468
Thr Glu Ile Phe Asn Met Pro Ala Ile Thr Gly Ala Gly Leu Met Thr
2270 2275 2280

tgt aga agc cag gcg gtg caa gaa cac gcc agc acc aac atg ggg cta 7516
Cys Arg Ser Gln Ala Val Gln Glu His Ala Ser Thr Asn Met Gly Leu
2285 2290 2295

gag gcc att att aga aag gca ctc atg ggt aaa tat gat cag tgg gaa 7564
Glu Ala Ile Ile Arg Lys Ala Leu Met Gly Lys Tyr Asp Gln Trp Glu
2300 2305 2310

gag ccc ccg ccg ctc ggc gcc aat gct ttt aac cct ctg aat gcc agc 7612
Glu Pro Pro Pro Leu Gly Ala Asn Ala Phe Asn Pro Leu Asn Ala Ser
2315 2320 2325

gcc agt ctg ccc gct gct gct atg ccc ata acc act gct gac gga cgg 7660
Ala Ser Leu Pro Ala Ala Ala Met Pro Ile Thr Thr Ala Asp Gly Arg
2330 2335 2340 2345

agt gac cac gca ctc acc tcg cca ggt gga ggt ggg aaa gcc aag gtc 7708
Ser Asp His Ala Leu Thr Ser Pro Gly Gly Gly Gly Lys Ala Lys Val
2350 2355 2360

tct ggc aga cct agc agc cga aaa gcc aag tcg cca gca cca ggc cta 7756
Ser Gly Arg Pro Ser Ser Arg Lys Ala Lys Ser Pro Ala Pro Gly Leu
2365 2370 2375

gcg tcc gga gac cga ccc cct tct gtc tcc tca gta cac tca gag ggg 7804
Ala Ser Gly Asp Arg Pro Pro Ser Val Ser Ser Val His Ser Glu Gly
2380 2385 2390

gac tgc aat cgc cga aca cca ctc acc aac cgt gtg tgg gag gac cgg 7852
Asp Cys Asn Arg Arg Thr Pro Leu Thr Asn Arg Val Trp Glu Asp Arg
2395 2400 2405

ccc tca tct gca ggg tcc acg cca ttc ccc tac aac cct ttg att atg 7900
Pro Ser Ser Ala Gly Ser Thr Pro Phe Pro Tyr Asn Pro Leu Ile Met
2410 2415 2420 2425

agg cta cag gca ggt gtc atg gcc tcc ccg ccc cca cct ggc ctt gcg 7948
Arg Leu Gln Ala Gly Val Met Ala Ser Pro Pro Pro Pro Gly Leu Ala
2430 2435 2440

gca ggc agc ggg ccc cta gct ggt ccc cac cac gcc tgg gat gag gag 7996
Ala Gly Ser Gly Pro Leu Ala Gly Pro His His Ala Trp Asp Glu Glu
2445 2450 2455

```
ccc aag cca ctg ctg tgt tca cag tat gag aca ctc tcg gac agc gag      8044
Pro Lys Pro Leu Leu Cys Ser Gln Tyr Glu Thr Leu Ser Asp Ser Glu
        2460                2465                2470

tga ccacggattg ggggggagcg gtgccaggtc ccgcacaagg cagaagcagc           8097
 *


ccagcatgga gcagacagct gctgactccc gagactgagg aaggagcccc tgagtctgcc    8157
tgcgcgtcca tccgtncgtc gtncactcat ctgtccatcc agagctggca ttctgcctgt    8217
ctaaagcctt aactaagact tccaccccgg gctggccctg cgcagtgacc ttacactcag    8277
gggattgttt accttggtgc tcganaaggg ggagtggaca ggaaggggag ggacaagccg    8337
ggccangagg gggggggaca ancaattcgt gtgtcaagtc gcactcntgc t             8388

<210> 7
<211> 2473
<212> PRT
<213> Mus musculus

<400> 7
Met Ser Gly Ser Thr Gln Pro Val Ala Gln Thr Trp Arg Ala Ala Glu
 1               5                  10                  15
Pro Arg Tyr Pro Pro His Gly Ile Ser Tyr Pro Val Gln Ile Ala Arg
            20                  25                  30
Ser His Thr Asp Val Gly Leu Leu Glu Tyr Gln His His Pro Arg Asp
        35                  40                  45
Tyr Thr Ser His Leu Ser Pro Gly Ser Ile Ile Gln Pro Gln Arg Arg
    50                  55                  60
Arg Pro Ser Leu Leu Ser Glu Phe Gln Pro Gly Ser Glu Arg Ser Gln
65                  70                  75                  80
Glu Leu His Leu Arg Pro Glu Ser Arg Thr Phe Leu Pro Glu Leu Gly
                85                  90                  95
Lys Pro Asp Ile Glu Phe Thr Glu Ser Lys Arg Pro Arg Leu Glu Leu
            100                 105                 110
Leu Pro Asp Thr Leu Leu Arg Pro Ser Pro Leu Leu Ala Thr Gly Gln
        115                 120                 125
Pro Ser Gly Ser Glu Asp Leu Thr Lys Asp Arg Ser Leu Ala Gly Lys
    130                 135                 140
Leu Glu Pro Val Ser Pro Pro Ser Pro Pro His Ala Asp Pro Glu Leu
145                 150                 155                 160
Glu Leu Ala Pro Ser Arg Leu Ser Lys Glu Glu Leu Ile Gln Asn Met
                165                 170                 175
Asp Arg Val Asp Arg Glu Ile Thr Met Val Glu Gln Gln Ile Ser Lys
            180                 185                 190
Leu Lys Lys Lys Gln Gln Gln Leu Glu Glu Glu Ala Ala Lys Pro Pro
        195                 200                 205
Glu Pro Glu Lys Pro Val Ser Pro Pro Pro Ile Glu Ser Lys His Arg
    210                 215                 220
Ser Leu Val Gln Ile Ile Tyr Asp Glu Asn Arg Lys Lys Ala Glu Ala
225                 230                 235                 240
Ala His Arg Ile Leu Glu Gly Leu Gly Pro Gln Val Glu Leu Pro Leu
                245                 250                 255
Tyr Asn Gln Pro Ser Asp Thr Arg Gln Tyr His Glu Asn Ile Lys Ile
            260                 265                 270
Asn Gln Ala Met Arg Lys Lys Leu Ile Leu Tyr Phe Lys Arg Arg Asn
        275                 280                 285
His Ala Arg Lys Gln Trp Glu Gln Arg Phe Cys Gln Arg Tyr Asp Gln
    290                 295                 300
Leu Met Glu Ala Trp Glu Lys Lys Val Glu Arg Ile Glu Asn Asn Pro
```

305 310 315 320
Arg Arg Arg Ala Lys Glu Ser Lys Val Arg Glu Tyr Tyr Glu Lys Gln
325 330 335
Phe Pro Glu Ile Arg Lys Gln Arg Glu Leu Gln Glu Arg Met Gln Ser
340 345 350
Arg Val Gly Gln Arg Gly Ser Gly Leu Ser Met Ser Ala Ala Arg Ser
355 360 365
Glu His Glu Val Ser Glu Ile Ile Asp Gly Leu Ser Glu Gln Glu Asn
370 375 380
Leu Glu Lys Gln Met Arg Gln Leu Ala Val Ile Pro Pro Met Leu Tyr
385 390 395 400
Asp Ala Asp Gln Gln Arg Ile Lys Phe Ile Asn Met Asn Gly Leu Met
405 410 415
Asp Asp Pro Met Lys Val Tyr Lys Asp Arg Gln Val Thr Asn Met Trp
420 425 430
Ser Glu Gln Glu Arg Asp Thr Phe Arg Glu Lys Phe Met Gln His Pro
435 440 445
Lys Asn Phe Gly Leu Ile Ala Ser Phe Leu Glu Arg Lys Thr Val Ala
450 455 460
Glu Cys Val Leu Tyr Tyr Tyr Leu Thr Lys Lys Asn Glu Asn Tyr Lys
465 470 475 480
Ser Leu Val Arg Arg Ser Tyr Arg Arg Arg Gly Lys Ser Gln Gln Gln
485 490 495
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Met Ala Arg Ser Ser
500 505 510
Gln Glu Glu Lys Glu Glu Lys Glu Lys Glu Lys Glu Ala Asp Lys Glu
515 520 525
Glu Glu Lys Gln Asp Ala Glu Asn Glu Lys Glu Glu Leu Ser Lys Glu
530 535 540
Lys Thr Asp Asp Thr Ser Gly Glu Asp Asn His Glu Lys Glu Ala Val
545 550 555 560
Ala Ser Lys Gly Arg Lys Thr Ala Asn Ser Gln Gly Arg Arg Lys Gly
565 570 575
Arg Ile Thr Arg Ser Met Ala Asn Glu Ala Asn His Glu Glu Thr Ala
580 585 590
Thr Pro Gln Gln Ser Ser Glu Leu Ala Ser Met Glu Met Asn Glu Ser
595 600 605
Ser Arg Trp Thr Glu Glu Glu Met Glu Thr Ala Lys Lys Gly Leu Leu
610 615 620
Glu His Gly Arg Asn Trp Ser Ala Ile Ala Arg Met Val Gly Ser Lys
625 630 635 640
Thr Val Ser Gln Cys Lys Asn Phe Tyr Phe Asn Tyr Lys Lys Arg Gln
645 650 655
Asn Leu Asp Glu Ile Leu Gln Gln His Lys Leu Lys Met Glu Lys Glu
660 665 670
Arg Asn Ala Arg Arg Lys Lys Lys Lys Thr Pro Ala Ala Ala Ser Glu
675 680 685
Glu Thr Ala Phe Pro Pro Ala Ala Glu Asp Glu Glu Met Glu Ala Ser
690 695 700
Gly Ala Ser Ala Asn Glu Glu Glu Leu Ala Glu Glu Ala Glu Ala Ser
705 710 715 720
Gln Ala Ser Gly Asn Glu Val Pro Arg Val Gly Glu Cys Ser Gly Pro
725 730 735
Ala Ala Val Asn Asn Ser Ser Asp Thr Glu Ser Val Pro Ser Pro Arg
740 745 750
Ser Glu Ala Met Lys Asp Thr Gly Pro Lys Pro Thr Gly Thr Glu Ala
755 760 765
Leu Pro Ala Ala Thr Gln Pro Pro Val Pro Pro Pro Glu Glu Pro Ala
770 775 780
Val Ala Pro Ala Glu Pro Ser Pro Val Pro Asp Ala Ser Gly Pro Pro

785 790 795 800
Ser Pro Glu Pro Ser His His Leu Pro His Pro Arg Leu Leu Trp Thr
805 810 815
Arg Met Asn Lys Lys Pro Arg Leu Leu Gln Leu Pro Arg Gln Arg Met
820 825 830
Pro Arg Ser Arg Ser Leu Arg Pro Arg Arg Ser Met Trp Glu Lys Pro
835 840 845
Glu Glu Pro Glu Ala Ser Glu Glu Pro Pro Glu Ser Val Lys Ser Asp
850 855 860
His Lys Glu Glu Thr Glu Glu Glu Pro Glu Asp Lys Ala Lys Gly Thr
865 870 875 880
Glu Ala Ile Glu Thr Val Ser Glu Ala Pro Leu Lys Val Glu Glu Ala
885 890 895
Gly Ser Lys Ala Ala Val Thr Lys Gly Ser Ser Ser Gly Ala Thr Gln
900 905 910
Asp Ser Asp Phe Ser Ala Thr Cys Ser Ala Asp Glu Val Asp Glu Pro
915 920 925
Glu Gly Gly Asp Lys Gly Arg Leu Leu Ser Pro Arg Pro Ser Leu Leu
930 935 940
Thr Pro Ala Gly Asp Pro Arg Ala Ser Thr Ser Pro Gln Lys Pro Leu
945 950 955 960
Asp Leu Lys Gln Leu Lys Gln Arg Ala Ala Ala Ile Pro Pro Ile Gln
965 970 975
Val Thr Lys Val His Glu Pro Pro Arg Glu Asp Thr Val Pro Pro Lys
980 985 990
Pro Val Pro Pro Val Pro Pro Pro Thr Gln His Leu Gln Pro Glu Gly
995 1000 1005
Asp Val Ser Gln Gln Ser Gly Gly Ser Pro Arg Gly Lys Ser Arg Ser
1010 1015 1020
Pro Val Pro Pro Ala Glu Lys Glu Ala Glu Lys Pro Ala Phe Phe Pro
1025 1030 1035 1040
Ala Phe Pro Thr Glu Gly Pro Lys Leu Pro Thr Glu Pro Pro Arg Trp
1045 1050 1055
Ser Ser Gly Leu Pro Phe Pro Ile Pro Pro Arg Glu Val Ile Lys Thr
1060 1065 1070
Ser Pro His Ala Ala Asp Pro Ser Ala Phe Ser Tyr Thr Pro Pro Gly
1075 1080 1085
His Pro Leu Pro Leu Gly Leu His Asp Ser Ala Arg Pro Val Leu Pro
1090 1095 1100
Arg Pro Pro Ile Ser Asn Pro Pro Pro Leu Ile Ser Ser Ala Lys His
1105 1110 1115 1120
Pro Gly Val Leu Glu Arg Gln Leu Gly Ala Ile Ser Gln Gln Gly Met
1125 1130 1135
Ser Val Gln Leu Arg Val Pro His Ser Glu His Ala Lys Ala Pro Met
1140 1145 1150
Gly Pro Leu Thr Met Gly Leu Pro Leu Ala Val Asp Pro Lys Lys Leu
1155 1160 1165
Gly Thr Ala Leu Gly Ser Ala Thr Ser Gly Ser Ile Thr Lys Gly Leu
1170 1175 1180
Pro Ser Thr Arg Ala Ala Asp Gly Pro Ser Tyr Arg Gly Ser Ile Thr
1185 1190 1195 1200
His Gly Thr Pro Ala Asp Val Leu Tyr Lys Gly Thr Ile Ser Arg Ile
1205 1210 1215
Val Gly Glu Asp Ser Pro Ser Arg Leu Asp Arg Ala Arg Glu Asp Thr
1220 1225 1230
Leu Pro Lys Gly His Val Ile Tyr Glu Gly Lys Lys Gly His Val Leu
1235 1240 1245
Ser Tyr Glu Gly Gly Met Ser Val Ser Gln Cys Ser Lys Glu Asp Gly
1250 1255 1260
Arg Ser Ser Ser Gly Pro Pro His Glu Thr Ala Ala Pro Lys Arg Thr

1265 1270 1275 1280
Tyr Asp Met Met Glu Gly Arg Val Gly Arg Thr Val Thr Ser Ala Ser
1285 1290 1295
Ile Glu Gly Leu Met Gly Arg Ala Ile Pro Glu Gln His Ser Pro His
1300 1305 1310
Leu Lys Glu Gln His His Ile Arg Gly Ser Ile Thr Gln Gly Ile Pro
1315 1320 1325
Arg Ser Tyr Val Glu Ala Gln Glu Asp Tyr Leu Arg Arg Glu Ala Lys
1330 1335 1340
Leu Leu Lys Arg Glu Gly Thr Pro Pro Pro Pro Pro Pro Pro Arg Asp
1345 1350 1355 1360
Leu Thr Glu Thr Tyr Lys Pro Arg Pro Leu Asp Pro Leu Gly Pro Leu
1365 1370 1375
Lys Leu Lys Pro Thr His Glu Gly Val Val Ala Thr Val Lys Glu Ala
1380 1385 1390
Gly Arg Ser Ile His Glu Ile Pro Arg Glu Glu Leu Arg Arg Thr Pro
1395 1400 1405
Glu Leu Pro Leu Ala Pro Arg Pro Leu Lys Glu Gly Ser Ile Thr Gln
1410 1415 1420
Gly Thr Pro Leu Lys Tyr Asp Ser Gly Ala Pro Ser Thr Gly Thr Lys
1425 1430 1435 1440
Lys His Asp Val Arg Ser Ile Ile Gly Ser Pro Gly Arg Pro Phe Pro
1445 1450 1455
Ala Leu His Pro Leu Asp Ile Met Ala Asp Ala Arg Ala Leu Glu Arg
1460 1465 1470
Ala Cys Tyr Glu Glu Ser Leu Lys Ser Arg Ser Gly Thr Ser Ser Gly
1475 1480 1485
Ala Gly Gly Ser Ile Thr Arg Gly Ala Pro Val Val Val Pro Glu Leu
1490 1495 1500
Gly Lys Pro Arg Gln Ser Pro Leu Thr Tyr Glu Asp His Gly Ala Pro
1505 1510 1515 1520
Phe Thr Ser His Leu Pro Arg Gly Ser Pro Val Thr Thr Arg Glu Pro
1525 1530 1535
Thr Pro Arg Leu Gln Glu Gly Ser Leu Leu Ser Ser Lys Ala Ser Gln
1540 1545 1550
Asp Arg Lys Leu Thr Ser Thr Pro Arg Glu Ile Ala Lys Ser Pro His
1555 1560 1565
Ser Thr Val Pro Glu His His Pro His Pro Ile Ser Pro Tyr Glu His
1570 1575 1580
Leu Leu Arg Gly Val Thr Gly Val Asp Leu Tyr Arg Gly His Ile Pro
1585 1590 1595 1600
Leu Ala Phe Asp Pro Thr Ser Ile Pro Arg Gly Ile Pro Leu Glu Ala
1605 1610 1615
Ala Ala Ala Ala Tyr Tyr Leu Pro Arg His Leu Ala Pro Ser Pro Thr
1620 1625 1630
Tyr Pro His Leu Tyr Pro Pro Tyr Leu Ile Arg Gly Tyr Pro Asp Thr
1635 1640 1645
Ala Ala Leu Glu Asn Arg Gln Thr Ile Ile Asn Asp Tyr Ile Thr Ser
1650 1655 1660
Gln Gln Met His His Asn Ala Ala Ser Ala Met Ala Gln Arg Ala Asp
1665 1670 1675 1680
Met Leu Arg Gly Leu Ser Pro Arg Glu Ser Ser Leu Ala Leu Asn Tyr
1685 1690 1695
Ala Ala Gly Pro Arg Gly Ile Ile Asp Leu Ser Gln Val Pro His Leu
1700 1705 1710
Pro Val Leu Val Pro Pro Thr Pro Gly Thr Pro Ala Thr Ala Ile Asp
1715 1720 1725
Arg Leu Ala Tyr Leu Pro Thr Ala Pro Pro Pro Phe Ser Ser Arg His
1730 1735 1740
Ser Ser Ser Pro Leu Ser Pro Gly Gly Pro Thr His Leu Ala Lys Pro

1745 1750 1755 1760
Thr Ala Thr Ser Ser Ser Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
1765 1770 1775
Asp Lys Ser Ile Leu Thr Ser Thr Thr Thr Val Glu His Ala Pro Ile
1780 1785 1790
Trp Arg Pro Gly Thr Glu Gln Ser Ser Gly Ala Gly Gly Ser Ser Arg
1795 1800 1805
Pro Ala Ser His Thr His Gln His Ser Pro Ile Ser Pro Arg Thr Gln
1810 1815 1820
Asp Ala Leu Gln Gln Arg Pro Ser Val Leu His Asn Thr Ser Met Lys
1825 1830 1835 1840
Gly Val Val Thr Ser Val Glu Pro Gly Thr Pro Thr Val Leu Arg Trp
1845 1850 1855
Ala Arg Ser Thr Ser Thr Ser Ser Pro Val Arg Pro Ala Ala Thr Phe
1860 1865 1870
Pro Pro Ala Thr His Cys Pro Leu Gly Gly Thr Leu Glu Gly Val Tyr
1875 1880 1885
Pro Thr Leu Met Glu Pro Val Leu Leu Pro Lys Glu Thr Ser Arg Val
1890 1895 1900
Ala Arg Pro Glu Arg Ala Arg Val Asp Ala Gly His Ala Phe Leu Thr
1905 1910 1915 1920
Lys Pro Pro Gly Arg Glu Pro Ala Ser Ser Pro Ser Lys Ser Ser Glu
1925 1930 1935
Pro Arg Ser Leu Ala Pro Pro Ser Ser Ser His Thr Ala Ile Ala Arg
1940 1945 1950
Thr Pro Ala Lys Asn Leu Ala Pro His His Ala Ser Pro Asp Pro Pro
1955 1960 1965
Ala Pro Thr Ser Ala Ser Asp Leu His Arg Glu Lys Thr Gln Ser Lys
1970 1975 1980
Pro Phe Ser Ile Gln Glu Leu Glu Leu Arg Ser Leu Gly Tyr His Ser
1985 1990 1995 2000
Gly Ala Gly Tyr Ser Pro Asp Gly Val Glu Pro Ile Ser Pro Val Ser
2005 2010 2015
Ser Pro Ser Leu Thr His Asp Lys Gly Leu Ser Lys Pro Leu Glu Glu
2020 2025 2030
Leu Glu Lys Ser His Leu Glu Gly Glu Leu Arg His Lys Gln Pro Gly
2035 2040 2045
Pro Met Lys Leu Ser Ala Glu Ala Ala His Leu Pro His Leu Arg Pro
2050 2055 2060
Leu Pro Glu Ser Gln Pro Ser Ser Ser Pro Leu Leu Gln Thr Ala Pro
2065 2070 2075 2080
Gly Ile Lys Gly His Gln Arg Val Val Thr Leu Ala Gln His Ile Ser
2085 2090 2095
Glu Val Ile Thr Gln Asp Tyr Thr Arg His His Pro Gln Gln Leu Ser
2100 2105 2110
Gly Pro Leu Pro Ala Pro Leu Tyr Ser Phe Pro Gly Ala Ser Cys Pro
2115 2120 2125
Val Leu Asp Leu Arg Arg Pro Pro Ser Asp Leu Tyr Leu Pro Pro Pro
2130 2135 2140
Asp His Gly Thr Pro Ala Arg Gly Ser Pro His Ser Glu Gly Gly Lys
2145 2150 2155 2160
Arg Ser Pro Glu Pro Ser Lys Thr Ser Val Leu Gly Ser Ser Glu Asp
2165 2170 2175
Ala Ile Glu Pro Val Ser Pro Pro Glu Gly Met Thr Glu Pro Gly His
2180 2185 2190
Ala Arg Ser Thr Ala Tyr Pro Leu Leu Tyr Arg Asp Gly Glu Gln Gly
2195 2200 2205
Glu Pro Arg Met Gly Leu Glu Ser Pro Gly Asn Thr Ser Gln Pro Pro
2210 2215 2220
Thr Phe Phe Ser Lys Leu Thr Glu Ser Asn Ser Ala Met Val Lys Ser

2225 2230 2235 2240
Lys Lys Gln Glu Ile Asn Lys Lys Leu Asn Thr His Asn Arg Asn Glu
2245 2250 2255
Pro Glu Tyr Asn Ile Gly Gln Pro Gly Thr Glu Ile Phe Asn Met Pro
2260 2265 2270
Ala Ile Thr Gly Ala Gly Leu Met Thr Cys Arg Ser Gln Ala Val Gln
2275 2280 2285
Glu His Ala Ser Thr Asn Met Gly Leu Glu Ala Ile Ile Arg Lys Ala
2290 2295 2300
Leu Met Gly Lys Tyr Asp Gln Trp Glu Glu Pro Pro Pro Leu Gly Ala
2305 2310 2315 2320
Asn Ala Phe Asn Pro Leu Asn Ala Ser Ala Ser Leu Pro Ala Ala Ala
2325 2330 2335
Met Pro Ile Thr Thr Ala Asp Gly Arg Ser Asp His Ala Leu Thr Ser
2340 2345 2350
Pro Gly Gly Gly Gly Lys Ala Lys Val Ser Gly Arg Pro Ser Ser Arg
2355 2360 2365
Lys Ala Lys Ser Pro Ala Pro Gly Leu Ala Ser Gly Asp Arg Pro Pro
2370 2375 2380
Ser Val Ser Ser Val His Ser Glu Gly Asp Cys Asn Arg Arg Thr Pro
2385 2390 2395 2400
Leu Thr Asn Arg Val Trp Glu Asp Arg Pro Ser Ser Ala Gly Ser Thr
2405 2410 2415
Pro Phe Pro Tyr Asn Pro Leu Ile Met Arg Leu Gln Ala Gly Val Met
2420 2425 2430
Ala Ser Pro Pro Pro Pro Gly Leu Ala Ala Gly Ser Gly Pro Leu Ala
2435 2440 2445
Gly Pro His His Ala Trp Asp Glu Glu Pro Lys Pro Leu Leu Cys Ser
2450 2455 2460
Gln Tyr Glu Thr Leu Ser Asp Ser Glu
2465 2470

<210> 8
<211> 7465
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (363)...(7124)

<221> misc_feature
<222> (1)...(7465)
<223> n = A,T,C or G

<400> 8
ggcacgaggg cagcgcaggc cgggccgcat ccccgtcccc gcgccagccg cccgcgcccg 60
ccatgcgcgc cccgcagcgg cccgcgcgtc cgggccccgc gtcgtagcgc ggcgggcgga 120
gaccgcaggc tctcagcccg gacccgccgc atcctcgagc ccgatcggcg ccgtagcccg 180
gcgccagcgc ccggtgccgc cgccggcgag tgctcctgag tctttgagga acacagcctc 240
ctggtggaag ttcgtggcac ctgtgacgag gtcacctgcc agcagatgac cgagaccagc 300
ccttagtcct aggtgtggtc aagagtgtct tggctccaa agcctacctg gaccctacca 360
cc atg tca gga tcc aca cag cct gtg gca cag aca tgg cgg gct gct 407
Met Ser Gly Ser Thr Gln Pro Val Ala Gln Thr Trp Arg Ala Ala
1 5 10 15

gag ccc cgc tac cca ccc cat ggc atc tcc tac ccg gtg cag ata gcc 455
Glu Pro Arg Tyr Pro Pro His Gly Ile Ser Tyr Pro Val Gln Ile Ala
20 25 30

```
cgg tcc cac acg cct ctg tac aac cag ccg tct gac aca cgc cag tac        503
Arg Ser His Thr Pro Leu Tyr Asn Gln Pro Ser Asp Thr Arg Gln Tyr
                 35                  40                  45

cat gaa aac atc aaa ata aac cag gcg atg cgg aag aag ctg atc ttg        551
His Glu Asn Ile Lys Ile Asn Gln Ala Met Arg Lys Lys Leu Ile Leu
             50                  55                  60

tac ttt aag cgg agg aac cac gcg cgc aag cag tgg gaa cag cgc ttc        599
Tyr Phe Lys Arg Arg Asn His Ala Arg Lys Gln Trp Glu Gln Arg Phe
         65                  70                  75

tgc cag cgc tat gac cag ctc atg gag gcg tgg gag aag aag gta gag        647
Cys Gln Arg Tyr Asp Gln Leu Met Glu Ala Trp Glu Lys Lys Val Glu
 80                  85                  90                  95

cgc ata gag aac aat ccg cga agg agg gcc aag gag agc aag gtg agg        695
Arg Ile Glu Asn Asn Pro Arg Arg Arg Ala Lys Glu Ser Lys Val Arg
                    100                 105                 110

gag tac tac gag aaa cag ttc ccg gag atc cgc aag cag cgg gag ctg        743
Glu Tyr Tyr Glu Lys Gln Phe Pro Glu Ile Arg Lys Gln Arg Glu Leu
                115                 120                 125

cag gag cgc atg cag agc agg gtg ggc cag cgt ggc agt ggg ctc tcc        791
Gln Glu Arg Met Gln Ser Arg Val Gly Gln Arg Gly Ser Gly Leu Ser
            130                 135                 140

atg tcg gct gcc cgc agt gag cat gag gtt tct gag atc att gat ggc        839
Met Ser Ala Ala Arg Ser Glu His Glu Val Ser Glu Ile Ile Asp Gly
        145                 150                 155

ttg tct gag cag gag aac ctg gag aag cag atg cgc cag ctg gcc gtg        887
Leu Ser Glu Gln Glu Asn Leu Glu Lys Gln Met Arg Gln Leu Ala Val
160                 165                 170                 175

atc ccg ccc atg ttg tac gac gcg gac cag cag agg atc aag ttc atc        935
Ile Pro Pro Met Leu Tyr Asp Ala Asp Gln Gln Arg Ile Lys Phe Ile
                180                 185                 190

aac atg aat gga ctc atg gat gac ccc atg aag gtc tac aag gac cgt        983
Asn Met Asn Gly Leu Met Asp Asp Pro Met Lys Val Tyr Lys Asp Arg
            195                 200                 205

cag gtt acc aac atg tgg agc gag cag gag agg gac acc ttc cgt gag       1031
Gln Val Thr Asn Met Trp Ser Glu Gln Glu Arg Asp Thr Phe Arg Glu
        210                 215                 220

aag ttt atg cag cac cct aag aac ttt ggc ctg att gcc tca ttc ctg       1079
Lys Phe Met Gln His Pro Lys Asn Phe Gly Leu Ile Ala Ser Phe Leu
    225                 230                 235

gag aga aag acg gtc gct gag tgt gtc ctc tat tac tac ctg acc aag       1127
Glu Arg Lys Thr Val Ala Glu Cys Val Leu Tyr Tyr Tyr Leu Thr Lys
240                 245                 250                 255

aag aat gaa aat tac aag agc ttg gtg agg cgg agc tat cgg cgc cgt       1175
Lys Asn Glu Asn Tyr Lys Ser Leu Val Arg Arg Ser Tyr Arg Arg Arg
                260                 265                 270
```

```
ggc aag agc cag cag cag cag cag cag caa caa cag cag cag cag cag      1223
Gly Lys Ser Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
            275                 280                 285

cag atg gca cgg agc agc cag gag gag aag gag gag aag gag aag gag      1271
Gln Met Ala Arg Ser Ser Gln Glu Glu Lys Glu Glu Lys Glu Lys Glu
        290                 295                 300

aag gag gcc gac aag gag gaa gag aag cag gat gcg gag aac gag aag      1319
Lys Glu Ala Asp Lys Glu Glu Glu Lys Gln Asp Ala Glu Asn Glu Lys
    305                 310                 315

gaa gaa ctc agc aag gag aag aca gac gac act tct ggc gag gac aac      1367
Glu Glu Leu Ser Lys Glu Lys Thr Asp Asp Thr Ser Gly Glu Asp Asn
320                 325                 330                 335

gat gag aaa gag gcc gtg gcc tcc aaa ggc cgc aaa act gcc aac agc      1415
Asp Glu Lys Glu Ala Val Ala Ser Lys Gly Arg Lys Thr Ala Asn Ser
                340                 345                 350

caa ggc cgc cgc aaa ggc cgt atc acg cgc tcc atg gcc aac gag gcc      1463
Gln Gly Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Ala Asn Glu Ala
            355                 360                 365

aac cat gag gag aca gcc acc cca cag caa agt tca gag ctg gct tcc      1511
Asn His Glu Glu Thr Ala Thr Pro Gln Gln Ser Ser Glu Leu Ala Ser
        370                 375                 380

atg gag atg aac gag agt tct cgc tgg act gag gaa gag atg gag aca      1559
Met Glu Met Asn Glu Ser Ser Arg Trp Thr Glu Glu Glu Met Glu Thr
    385                 390                 395

gca aag aaa ggc ctc ctg gaa cat ggg agg aac tgg tca gcc att gcc      1607
Ala Lys Lys Gly Leu Leu Glu His Gly Arg Asn Trp Ser Ala Ile Ala
400                 405                 410                 415

cgc atg gtg ggc tcc aag acc gtg tcc cag tgt aag aac ttc tac ttc      1655
Arg Met Val Gly Ser Lys Thr Val Ser Gln Cys Lys Asn Phe Tyr Phe
                420                 425                 430

aac tac aag aag agg cag aac ctg gac gaa atc ctt cag cag cac aag      1703
Asn Tyr Lys Lys Arg Gln Asn Leu Asp Glu Ile Leu Gln Gln His Lys
            435                 440                 445

cta aag atg gag aag gag agg aac gct cgg agg aag aag aag aag acc      1751
Leu Lys Met Glu Lys Glu Arg Asn Ala Arg Arg Lys Lys Lys Lys Thr
        450                 455                 460

cca gct gcg gcg agc gag gag aca gcc ttc cca cct gcc gct gag gac      1799
Pro Ala Ala Ala Ser Glu Glu Thr Ala Phe Pro Pro Ala Ala Glu Asp
    465                 470                 475

gaa gag atg gaa gca tca ggc gca agt gcc aat gag gaa gag ctg gcg      1847
Glu Glu Met Glu Ala Ser Gly Ala Ser Ala Asn Glu Glu Glu Leu Ala
480                 485                 490                 495

gag gag gca gaa gcc tca cag gcc tct ggg aat gag gtt ccc aga gtt      1895
Glu Glu Ala Glu Ala Ser Gln Ala Ser Gly Asn Glu Val Pro Arg Val
                500                 505                 510
```

```
ggg gag tgc agt ggc cca gct gct gtc aac aac agc tct gat act gag      1943
Gly Glu Cys Ser Gly Pro Ala Ala Val Asn Asn Ser Ser Asp Thr Glu
            515                 520                 525

agt gtc cca tcc ccg cgt tca gaa gcc acg aag gac act ggg cct aaa      1991
Ser Val Pro Ser Pro Arg Ser Glu Ala Thr Lys Asp Thr Gly Pro Lys
        530                 535                 540

ccc act ggc act gaa gca ttg ccc gct gcc acc cag cca cct gtt cct      2039
Pro Thr Gly Thr Glu Ala Leu Pro Ala Ala Thr Gln Pro Pro Val Pro
    545                 550                 555

cct cca gaa gaa ccg gca gta gcc cct gct gag ccc tcc cca gtc cct      2087
Pro Pro Glu Glu Pro Ala Val Ala Pro Ala Glu Pro Ser Pro Val Pro
560                 565                 570                 575

gat gcc agt ggc cca cca tcc cca gag cct tcc cat cac ctg ccg cac      2135
Asp Ala Ser Gly Pro Pro Ser Pro Glu Pro Ser His His Leu Pro His
                580                 585                 590

ccc cgg cta ctg tgg aca agg atg aac aag aag ccc cgg ctg ctc cag      2183
Pro Arg Leu Leu Trp Thr Arg Met Asn Lys Lys Pro Arg Leu Leu Gln
            595                 600                 605

ctc ccc aga cag agg atg cca agg agc aga agt ctg agg ccg agg aga      2231
Leu Pro Arg Gln Arg Met Pro Arg Ser Arg Ser Leu Arg Pro Arg Arg
        610                 615                 620

tcg atg tgg gaa aag cca gag gag ccc gag gct tct gag aag ccc ccg      2279
Ser Met Trp Glu Lys Pro Glu Glu Pro Glu Ala Ser Glu Lys Pro Pro
    625                 630                 635

aag agt gta aag agt gac cac aag aag gag acc gag gaa gag cct gaa      2327
Lys Ser Val Lys Ser Asp His Lys Lys Glu Thr Glu Glu Glu Pro Glu
640                 645                 650                 655

gac aaa gcc aag ggc aca gag gcc att gaa act gtg tct gag gca cca      2375
Asp Lys Ala Lys Gly Thr Glu Ala Ile Glu Thr Val Ser Glu Ala Pro
                660                 665                 670

ctt aag gtg gag aag gct ggt agc aag gca gct gtg acc aag ggt tcc      2423
Leu Lys Val Glu Lys Ala Gly Ser Lys Ala Ala Val Thr Lys Gly Ser
            675                 680                 685

agc tca ggt gcc acc cag gac agt gac tcc agt gcc acc tgc agt gcc      2471
Ser Ser Gly Ala Thr Gln Asp Ser Asp Ser Ser Ala Thr Cys Ser Ala
        690                 695                 700

gat gag gtg gac gaa ccc gaa gga ggt gac aag ggc agg ctg ctg tca      2519
Asp Glu Val Asp Glu Pro Glu Gly Gly Asp Lys Gly Arg Leu Leu Ser
    705                 710                 715

cca agg ccc agc ctc ctc acc ccg gct gga gat ccc cgg gcc agt acc      2567
Pro Arg Pro Ser Leu Leu Thr Pro Ala Gly Asp Pro Arg Ala Ser Thr
720                 725                 730                 735

tcg ccc cag aag ccg ctg gac ctg aag cag ctg aag cag cga gca gcc      2615
Ser Pro Gln Lys Pro Leu Asp Leu Lys Gln Leu Lys Gln Arg Ala Ala
                740                 745                 750
```

```
gcc atc ccc cct atc gtc acc aag gtc cat gag ccc ccc cgg gag gac      2663
Ala Ile Pro Pro Ile Val Thr Lys Val His Glu Pro Pro Arg Glu Asp
            755                 760                 765

aca gta ccc cca aag cca gtt ccc cct gtg cct cca ccc acg cag cac      2711
Thr Val Pro Pro Lys Pro Val Pro Pro Val Pro Pro Pro Thr Gln His
        770                 775                 780

cta cag cca gag ggt gac gtg tct cag cag tcg gga gga agt cca cgt      2759
Leu Gln Pro Glu Gly Asp Val Ser Gln Gln Ser Gly Gly Ser Pro Arg
    785                 790                 795

ggc aag tcc cgc agc cca gtg cct cct gcc gag aaa gag gca gag aaa      2807
Gly Lys Ser Arg Ser Pro Val Pro Pro Ala Glu Lys Glu Ala Glu Lys
800                 805                 810                 815

ccc gca ttc ttt ccg gct ttc cca act gag ggc cca aag cta ccg act      2855
Pro Ala Phe Phe Pro Ala Phe Pro Thr Glu Gly Pro Lys Leu Pro Thr
                820                 825                 830

gag ccc cca cgc tgg tca tcg ggc ctg ccc ttc ccc atc cct cca cgg      2903
Glu Pro Pro Arg Trp Ser Ser Gly Leu Pro Phe Pro Ile Pro Pro Arg
            835                 840                 845

gag gtg atc aag act tcc cca cac gcc gct gac ccc tct gcc ttc tcc      2951
Glu Val Ile Lys Thr Ser Pro His Ala Ala Asp Pro Ser Ala Phe Ser
        850                 855                 860

tac aca ccc ccc ggt cac ccg ctg cct ctg ggc ctc cac gat agt gcc      2999
Tyr Thr Pro Pro Gly His Pro Leu Pro Leu Gly Leu His Asp Ser Ala
    865                 870                 875

cgg ccc gtc ctg cca cgt ccc ccc atc tct aac ccc cca ccc ctc atc      3047
Arg Pro Val Leu Pro Arg Pro Pro Ile Ser Asn Pro Pro Pro Leu Ile
880                 885                 890                 895

tcc tct gcc aag cat ccc ggc gta ctt gag agg cag ctg ggt gcc atc      3095
Ser Ser Ala Lys His Pro Gly Val Leu Glu Arg Gln Leu Gly Ala Ile
                900                 905                 910

tcc cag cag ggg atg tca gtc cag ctt cgt gtg cct cac tca gag cat      3143
Ser Gln Gln Gly Met Ser Val Gln Leu Arg Val Pro His Ser Glu His
            915                 920                 925

gcc aag gcc ccc atg ggc cct ctc acc atg ggg ctg ccc ctt gcc gtg      3191
Ala Lys Ala Pro Met Gly Pro Leu Thr Met Gly Leu Pro Leu Ala Val
        930                 935                 940

gac cct aag aag ctg ggg aca gca ctg ggc tcc gcc acc agt gga agc      3239
Asp Pro Lys Lys Leu Gly Thr Ala Leu Gly Ser Ala Thr Ser Gly Ser
    945                 950                 955

atc acc aag ggc ctc ccc agt acc cgg gct gca gac ggc ccc agc tac      3287
Ile Thr Lys Gly Leu Pro Ser Thr Arg Ala Ala Asp Gly Pro Ser Tyr
960                 965                 970                 975

aga ggc tct atc acc cac ggc acg ccc gca gac gtc ctc tac aag ggt      3335
Arg Gly Ser Ile Thr His Gly Thr Pro Ala Asp Val Leu Tyr Lys Gly
                980                 985                 990
```

```
acc atc agc agg atc gtc ggt gag gac agc cca agt cgc ctt gac cgg      3383
Thr Ile Ser Arg Ile Val Gly Glu Asp Ser Pro Ser Arg Leu Asp Arg
                995                     1000                1005

gca cga gag gac acc ctg ccc aag ggc cat gtc atc tat gag ggc aag      3431
Ala Arg Glu Asp Thr Leu Pro Lys Gly His Val Ile Tyr Glu Gly Lys
        1010                1015                1020

aaa ggc cac gtc cta tcc tat gaa ggt ggt atg tcc gtg tca cag tgc      3479
Lys Gly His Val Leu Ser Tyr Glu Gly Gly Met Ser Val Ser Gln Cys
    1025                1030                1035

tct aag gag gat gga agg agc agc tcg ggc cca ccc cat gag act gcc      3527
Ser Lys Glu Asp Gly Arg Ser Ser Ser Gly Pro Pro His Glu Thr Ala
1040                1045                1050                1055

gcc cct aaa cgc acc tat gac atg atg gag ggc cgt gta ggc agg act      3575
Ala Pro Lys Arg Thr Tyr Asp Met Met Glu Gly Arg Val Gly Arg Thr
                1060                1065                1070

gtc acc tca gcc agc ata gag gga ctc atg ggc cgc gcc atc cct gag      3623
Val Thr Ser Ala Ser Ile Glu Gly Leu Met Gly Arg Ala Ile Pro Glu
            1075                1080                1085

cag cac agc ccc cac ctc aag gag cag cat cac atc cga ggc tcc atc      3671
Gln His Ser Pro His Leu Lys Glu Gln His His Ile Arg Gly Ser Ile
        1090                1095                1100

acg caa ggc atc ccg agg tcc tat gtg gag gcg cag gag gac tac tta      3719
Thr Gln Gly Ile Pro Arg Ser Tyr Val Glu Ala Gln Glu Asp Tyr Leu
    1105                1110                1115

cgg cgg gag gcc aag ctc ttg aag cga gaa ggg aca cca cca ccc cca      3767
Arg Arg Glu Ala Lys Leu Leu Lys Arg Glu Gly Thr Pro Pro Pro Pro
1120                1125                1130                1135

cca cca cct cgg gac ctg act gag acc tac aag ccc cgg ccc ctg gac      3815
Pro Pro Pro Arg Asp Leu Thr Glu Thr Tyr Lys Pro Arg Pro Leu Asp
                1140                1145                1150

cct ctg ggt ccc ctg aag ctg aag ccg act cac gag ggt gtg gta gca      3863
Pro Leu Gly Pro Leu Lys Leu Lys Pro Thr His Glu Gly Val Val Ala
            1155                1160                1165

act gtg aag gag gcg ggc cgc tct atc cat gag atc ccg aga gag gag      3911
Thr Val Lys Glu Ala Gly Arg Ser Ile His Glu Ile Pro Arg Glu Glu
        1170                1175                1180

ctg cgc cgc aca cct gag cta ccc ctg gca cca cgg cct ctg aag gag      3959
Leu Arg Arg Thr Pro Glu Leu Pro Leu Ala Pro Arg Pro Leu Lys Glu
    1185                1190                1195

ggt tcc atc acc cag ggc acc cca ctc aag tac gac tct ggg gca ccc      4007
Gly Ser Ile Thr Gln Gly Thr Pro Leu Lys Tyr Asp Ser Gly Ala Pro
1200                1205                1210                1215

tcc act ggc acc aag aaa cac gac gtg cgc tcc atc atc ggc agc ccc      4055
Ser Thr Gly Thr Lys Lys His Asp Val Arg Ser Ile Ile Gly Ser Pro
                1220                1225                1230
```

```
ggc cgg cct ttc cct gcc ctg cac ccg ctg gac ata atg gct gac gcc      4103
Gly Arg Pro Phe Pro Ala Leu His Pro Leu Asp Ile Met Ala Asp Ala
            1235                1240                1245

cgg gca ctg gag cgt gcc tgc tat gaa gag agt ctg aag agc cgg tca      4151
Arg Ala Leu Glu Arg Ala Cys Tyr Glu Glu Ser Leu Lys Ser Arg Ser
        1250                1255                1260

ggg acc agc agt ggt gca ggg ggc tcc atc aca cgt ggg gct cca gtc      4199
Gly Thr Ser Ser Gly Ala Gly Gly Ser Ile Thr Arg Gly Ala Pro Val
    1265                1270                1275

gtc gtg cct gaa ctg ggc aag cca cgg caa agc cca ctg act tac gaa      4247
Val Val Pro Glu Leu Gly Lys Pro Arg Gln Ser Pro Leu Thr Tyr Glu
1280                1285                1290                1295

gac cac ggg gca ccc ttc acc agt cac ctg cca cgt ggc tcc cct gtg      4295
Asp His Gly Ala Pro Phe Thr Ser His Leu Pro Arg Gly Ser Pro Val
                1300                1305                1310

acc acg agg gag ccc acg cca cgc ctt cag gaa ggc agc ctc cta tcc      4343
Thr Thr Arg Glu Pro Thr Pro Arg Leu Gln Glu Gly Ser Leu Leu Ser
            1315                1320                1325

agc aag gcg tcc cag gac cgg aag ctg aca tct aca ccc cgg gag atc      4391
Ser Lys Ala Ser Gln Asp Arg Lys Leu Thr Ser Thr Pro Arg Glu Ile
        1330                1335                1340

gcc aag tcc cca cac agc act gtg ccc gag cac cac cct cac ccc atc      4439
Ala Lys Ser Pro His Ser Thr Val Pro Glu His His Pro His Pro Ile
    1345                1350                1355

tcc ccc tat gag cac ttg ctc cgg ggc gtg act ggt gtg gac ctg tac      4487
Ser Pro Tyr Glu His Leu Leu Arg Gly Val Thr Gly Val Asp Leu Tyr
1360                1365                1370                1375

cgt ggt cac atc cca ttg gcc ttt gac ccc acc tcc ata ccc cga ggg      4535
Arg Gly His Ile Pro Leu Ala Phe Asp Pro Thr Ser Ile Pro Arg Gly
                1380                1385                1390

atc cct ctg gaa gca gca gcc gca gcc tac tac ctg ccc cgg cac ttg      4583
Ile Pro Leu Glu Ala Ala Ala Ala Ala Tyr Tyr Leu Pro Arg His Leu
            1395                1400                1405

gcc ccc agc ccc acc tac cca cac ctg tac cca cct tac ctc atc cgc      4631
Ala Pro Ser Pro Thr Tyr Pro His Leu Tyr Pro Pro Tyr Leu Ile Arg
        1410                1415                1420

ggc tac cct gac acg gcg gcc ctg gag aac cgc cag acc atc atc aat      4679
Gly Tyr Pro Asp Thr Ala Ala Leu Glu Asn Arg Gln Thr Ile Ile Asn
    1425                1430                1435

gac tac atc acc tcg cag cag atg cac cac aac gct gcc tcc gcc atg      4727
Asp Tyr Ile Thr Ser Gln Gln Met His His Asn Ala Ala Ser Ala Met
1440                1445                1450                1455

gcc cag cgt gct gac atg ctg agg ggt ctg tca ccg cga gag tcc tcg      4775
Ala Gln Arg Ala Asp Met Leu Arg Gly Leu Ser Pro Arg Glu Ser Ser
                1460                1465                1470
```

```
ctg gcc ctc aat tat gcc gct ggc cca aga ggc att atc gac ctg tcc      4823
Leu Ala Leu Asn Tyr Ala Ala Gly Pro Arg Gly Ile Ile Asp Leu Ser
                1475                1480                1485

caa gtg cca cac ctg ccc gtg ctg gtg cca cca acg cca ggc acc cct      4871
Gln Val Pro His Leu Pro Val Leu Val Pro Pro Thr Pro Gly Thr Pro
        1490                1495                1500

gcc acc gcc atc gac cgc ctt gcc tac ctc ccc act gcg ccc cca ccc      4919
Ala Thr Ala Ile Asp Arg Leu Ala Tyr Leu Pro Thr Ala Pro Pro Pro
    1505                1510                1515

ttc agc agc cgc cac agt agc tca ccg ctg tcc cca gga ggc ccc act      4967
Phe Ser Ser Arg His Ser Ser Ser Pro Leu Ser Pro Gly Gly Pro Thr
1520                1525                1530                1535

cac cta gct aaa cca act gcc aca tct tca tcg gag cgg gaa cgg gaa      5015
His Leu Ala Lys Pro Thr Ala Thr Ser Ser Ser Glu Arg Glu Arg Glu
                1540                1545                1550

cgt gag cgg gaa cga gac aag tcc atc ctc acg tct acc act aca gtg      5063
Arg Glu Arg Glu Arg Asp Lys Ser Ile Leu Thr Ser Thr Thr Thr Val
                1555                1560                1565

gag cat gca ccc atc tgg aga cct ggt acg gag cag agc agc ggg gct      5111
Glu His Ala Pro Ile Trp Arg Pro Gly Thr Glu Gln Ser Ser Gly Ala
        1570                1575                1580

ggg ggc agc agc cgc ccc gcc tcc cac acc cac cag cac tcg ccc atc      5159
Gly Gly Ser Ser Arg Pro Ala Ser His Thr His Gln His Ser Pro Ile
    1585                1590                1595

tcc ccc cgg acc cag gac gcc ttg cag cag agg ccc agt gtg ctg cac      5207
Ser Pro Arg Thr Gln Asp Ala Leu Gln Gln Arg Pro Ser Val Leu His
1600                1605                1610                1615

aac acg agc atg aag ggc gtg gtc acc tcc gtg gaa ccc ggc acg ccc      5255
Asn Thr Ser Met Lys Gly Val Val Thr Ser Val Glu Pro Gly Thr Pro
                1620                1625                1630

acg gtc ctg agg tgg gcc agg tcc acc tcc acc tct tcg cct gtc cgc      5303
Thr Val Leu Arg Trp Ala Arg Ser Thr Ser Thr Ser Ser Pro Val Arg
                1635                1640                1645

cca gct gcc aca ttc cca cct gcc acc cac tgc cca ctt ggt ggc acc      5351
Pro Ala Ala Thr Phe Pro Pro Ala Thr His Cys Pro Leu Gly Gly Thr
        1650                1655                1660

ctt gaa ggg gtc tac cct acc ctc atg gag ccc gtc ctg tta ccc aag      5399
Leu Glu Gly Val Tyr Pro Thr Leu Met Glu Pro Val Leu Leu Pro Lys
    1665                1670                1675

gag acc tct cgg gtc gcc cgg ccc gag cgg gcc cgg gtg gac gct ggc      5447
Glu Thr Ser Arg Val Ala Arg Pro Glu Arg Ala Arg Val Asp Ala Gly
1680                1685                1690                1695

cat gcc ttt ctt acc aaa ccc ccg ggc cgg gag ccc gcc tcc tca ccc      5495
His Ala Phe Leu Thr Lys Pro Pro Gly Arg Glu Pro Ala Ser Ser Pro
                1700                1705                1710
```

```
agc aag agc tcc gag ccc cga tcc cta gca ccc ccc agc tcc agc cac      5543
Ser Lys Ser Ser Glu Pro Arg Ser Leu Ala Pro Pro Ser Ser Ser His
            1715                1720                1725

aca gcc atc gcc cgc acc cca gca aag aac ctt gca ccc cac cat gcc      5591
Thr Ala Ile Ala Arg Thr Pro Ala Lys Asn Leu Ala Pro His His Ala
        1730                1735                1740

agt ccg gac ccg ccg gcg ccc acc tcg gcc tca gat ctg cac cga gaa      5639
Ser Pro Asp Pro Pro Ala Pro Thr Ser Ala Ser Asp Leu His Arg Glu
    1745                1750                1755

aag act caa agt aaa ccc ttt tcc atc cag gaa ttg gaa ctc cgt tct      5687
Lys Thr Gln Ser Lys Pro Phe Ser Ile Gln Glu Leu Glu Leu Arg Ser
1760                1765                1770                1775

ctg ggt tac cac agt gga gct ggc tac agc ccc gat ggg gtg gag ccc      5735
Leu Gly Tyr His Ser Gly Ala Gly Tyr Ser Pro Asp Gly Val Glu Pro
                1780                1785                1790

atc agc ccg gtg agc tcc ccc agc ctg acc cac gac aag ggg ctc tcc      5783
Ile Ser Pro Val Ser Ser Pro Ser Leu Thr His Asp Lys Gly Leu Ser
            1795                1800                1805

aaa cct ctg gaa gag cta gag aag agc cac ttg gaa ggg gag ctg cgg      5831
Lys Pro Leu Glu Glu Leu Glu Lys Ser His Leu Glu Gly Glu Leu Arg
        1810                1815                1820

cac aag cag cca ggc ccc atg aag ctc agc gcg gag gct gcc cat ctc      5879
His Lys Gln Pro Gly Pro Met Lys Leu Ser Ala Glu Ala Ala His Leu
    1825                1830                1835

cca cat ctg cgg cca ctg ccc gag agc cag ccc tca tcc agc cca ctc      5927
Pro His Leu Arg Pro Leu Pro Glu Ser Gln Pro Ser Ser Ser Pro Leu
1840                1845                1850                1855

ctc cag act gcc cca ggc atc aaa ggt cac cag agg gtg gtc acc ctg      5975
Leu Gln Thr Ala Pro Gly Ile Lys Gly His Gln Arg Val Val Thr Leu
                1860                1865                1870

gct cag cac atc agc gag gtc att acg cag gac tac acg cgc cac cac      6023
Ala Gln His Ile Ser Glu Val Ile Thr Gln Asp Tyr Thr Arg His His
            1875                1880                1885

ccg cag cag ctc agt ggc ccc ctt ccc gcc cct ctc tac tcc ttt ccc      6071
Pro Gln Gln Leu Ser Gly Pro Leu Pro Ala Pro Leu Tyr Ser Phe Pro
        1890                1895                1900

gga gcc agc tgc cct gtc ctg gat ctt cgc cgc cca ccc agt gac ctc      6119
Gly Ala Ser Cys Pro Val Leu Asp Leu Arg Arg Pro Pro Ser Asp Leu
    1905                1910                1915

tac ctc cca ccc ccc gac cat ggc acc cca gcc cgg gga tcc ccc cac      6167
Tyr Leu Pro Pro Pro Asp His Gly Thr Pro Ala Arg Gly Ser Pro His
1920                1925                1930                1935

agt gaa ggg ggc aaa agg tcc cca gaa ccc agc aaa aca tcg gtc ctg      6215
Ser Glu Gly Gly Lys Arg Ser Pro Glu Pro Ser Lys Thr Ser Val Leu
                1940                1945                1950
```

ggc agc agc gag gat gcc att gag cct gtg tcc cca cca gag ggc atg 6263
Gly Ser Ser Glu Asp Ala Ile Glu Pro Val Ser Pro Pro Glu Gly Met
1955 1960 1965

act gag cca gga cat gct cgg agc act gcg tac cca ctg ctg tat cga 6311
Thr Glu Pro Gly His Ala Arg Ser Thr Ala Tyr Pro Leu Leu Tyr Arg
1970 1975 1980

gac ggg gaa cag ggc gag ccc agg atg ggt cta gag tct cca ggc aac 6359
Asp Gly Glu Gln Gly Glu Pro Arg Met Gly Leu Glu Ser Pro Gly Asn
1985 1990 1995

acc agc cag ccg cca acc ttc ttc agt aag ctg act gag agc aac tcc 6407
Thr Ser Gln Pro Pro Thr Phe Phe Ser Lys Leu Thr Glu Ser Asn Ser
2000 2005 2010 2015

gcc atg gtg aag tcg aag aag cag gag atc aac aag aaa ctc aac acc 6455
Ala Met Val Lys Ser Lys Lys Gln Glu Ile Asn Lys Lys Leu Asn Thr
2020 2025 2030

cac aac cgg aac gag cca gaa tac aat att ggc cag cct ggg acg gaa 6503
His Asn Arg Asn Glu Pro Glu Tyr Asn Ile Gly Gln Pro Gly Thr Glu
2035 2040 2045

atc ttc aac atg ccc gcc atc act gga gca ggc ctt atg acc tgt aga 6551
Ile Phe Asn Met Pro Ala Ile Thr Gly Ala Gly Leu Met Thr Cys Arg
2050 2055 2060

agc cag gcg gtg caa gaa cac gcc agc acc aac atg ggg cta gag gcc 6599
Ser Gln Ala Val Gln Glu His Ala Ser Thr Asn Met Gly Leu Glu Ala
2065 2070 2075

att att aga aag gca ctc atg ggt aaa tat gat cag tgg gaa gag ccc 6647
Ile Ile Arg Lys Ala Leu Met Gly Lys Tyr Asp Gln Trp Glu Glu Pro
2080 2085 2090 2095

ccg ccg ctc ggc gcc aat gct ttt aac cct ctg aat gcc agc gcc agt 6695
Pro Pro Leu Gly Ala Asn Ala Phe Asn Pro Leu Asn Ala Ser Ala Ser
2100 2105 2110

ctg ccc gct gct gct atg ccc ata acc act gct gac gga cgg agt gac 6743
Leu Pro Ala Ala Ala Met Pro Ile Thr Thr Ala Asp Gly Arg Ser Asp
2115 2120 2125

cac gca ctc acc tcg cca ggt gga ggt ggg aaa gcc aag gtc tct ggc 6791
His Ala Leu Thr Ser Pro Gly Gly Gly Gly Lys Ala Lys Val Ser Gly
2130 2135 2140

aga cct agc agc cga aaa gcc aag tcg cca gca cca ggc cta gcg tcc 6839
Arg Pro Ser Ser Arg Lys Ala Lys Ser Pro Ala Pro Gly Leu Ala Ser
2145 2150 2155

gga gac cga ccc cct tct gtc tcc tca gta cac tca gag ggg gac tgc 6887
Gly Asp Arg Pro Pro Ser Val Ser Ser Val His Ser Glu Gly Asp Cys
2160 2165 2170 2175

aat cgc cga aca cca ctc acc aac cgt gtg tgg gag gac cgg ccc tca 6935
Asn Arg Arg Thr Pro Leu Thr Asn Arg Val Trp Glu Asp Arg Pro Ser
2180 2185 2190

tct gca ggg tcc acg cca ttc ccc tac aac cct ttg att atg agg cta 6983
Ser Ala Gly Ser Thr Pro Phe Pro Tyr Asn Pro Leu Ile Met Arg Leu
2195 2200 2205

cag gca ggt gtc atg gcc tcc ccg ccc cca cct ggc ctt gcg gca ggc 7031
Gln Ala Gly Val Met Ala Ser Pro Pro Pro Pro Gly Leu Ala Ala Gly
2210 2215 2220

agc ggg ccc cta gct ggt ccc cac cac gcc tgg gat gag gag ccc aag 7079
Ser Gly Pro Leu Ala Gly Pro His His Ala Trp Asp Glu Glu Pro Lys
2225 2230 2235

cca ctg ctg tgt tca cag tat gag aca ctc tcg gac agc gag tga 7124
Pro Leu Leu Cys Ser Gln Tyr Glu Thr Leu Ser Asp Ser Glu *
2240 2245 2250

ccacggattg gggggagcg gtgccaggtc ccgcacaagg cagaagcagc ccagcatgga 7184
gcagacagct gctgactccc gagactgagg aaggagcccc tgagtctgcc tgcgcgtcca 7244
tccgtncgtc gtncactcat ctgtccatcc agagctggca ttctgcctgt ctaaagcctt 7304
aactaagact tccaccccgg gctggccctg cgcagtgacc ttacactcag gggattgttt 7364
accttggtgc tcganaaggg ggagtggaca ggaaggggag ggacaagccg ggccangagg 7424
ggggggggaca ancaattcgt gtgtcaagtc gcactcntgc t 7465

<210> 9
<211> 2253
<212> PRT
<213> Mus musculus

<400> 9
Met Ser Gly Ser Thr Gln Pro Val Ala Gln Thr Trp Arg Ala Ala Glu
1 5 10 15
Pro Arg Tyr Pro Pro His Gly Ile Ser Tyr Pro Val Gln Ile Ala Arg
20 25 30
Ser His Thr Pro Leu Tyr Asn Gln Pro Ser Asp Thr Arg Gln Tyr His
35 40 45
Glu Asn Ile Lys Ile Asn Gln Ala Met Arg Lys Lys Leu Ile Leu Tyr
50 55 60
Phe Lys Arg Arg Asn His Ala Arg Lys Gln Trp Glu Gln Arg Phe Cys
65 70 75 80
Gln Arg Tyr Asp Gln Leu Met Glu Ala Trp Glu Lys Lys Val Glu Arg
85 90 95
Ile Glu Asn Asn Pro Arg Arg Arg Ala Lys Glu Ser Lys Val Arg Glu
100 105 110
Tyr Tyr Glu Lys Gln Phe Pro Glu Ile Arg Lys Gln Arg Glu Leu Gln
115 120 125
Glu Arg Met Gln Ser Arg Val Gly Gln Arg Gly Ser Gly Leu Ser Met
130 135 140
Ser Ala Ala Arg Ser Glu His Glu Val Ser Glu Ile Ile Asp Gly Leu
145 150 155 160
Ser Glu Gln Glu Asn Leu Glu Lys Gln Met Arg Gln Leu Ala Val Ile
165 170 175
Pro Pro Met Leu Tyr Asp Ala Asp Gln Gln Arg Ile Lys Phe Ile Asn
180 185 190
Met Asn Gly Leu Met Asp Asp Pro Met Lys Val Tyr Lys Asp Arg Gln
195 200 205
Val Thr Asn Met Trp Ser Glu Gln Glu Arg Asp Thr Phe Arg Glu Lys
210 215 220
Phe Met Gln His Pro Lys Asn Phe Gly Leu Ile Ala Ser Phe Leu Glu
225 230 235 240
Arg Lys Thr Val Ala Glu Cys Val Leu Tyr Tyr Tyr Leu Thr Lys Lys

245 250 255
Asn Glu Asn Tyr Lys Ser Leu Val Arg Arg Ser Tyr Arg Arg Arg Gly
260 265 270
Lys Ser Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
275 280 285
Met Ala Arg Ser Ser Gln Glu Glu Lys Glu Glu Lys Glu Lys Glu Lys
290 295 300
Glu Ala Asp Lys Glu Glu Glu Lys Gln Asp Ala Glu Asn Glu Lys Glu
305 310 315 320
Glu Leu Ser Lys Glu Lys Thr Asp Asp Thr Ser Gly Glu Asp Asn Asp
325 330 335
Glu Lys Glu Ala Val Ala Ser Lys Gly Arg Lys Thr Ala Asn Ser Gln
340 345 350
Gly Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Ala Asn Glu Ala Asn
355 360 365
His Glu Glu Thr Ala Thr Pro Gln Gln Ser Ser Glu Leu Ala Ser Met
370 375 380
Glu Met Asn Glu Ser Ser Arg Trp Thr Glu Glu Glu Met Glu Thr Ala
385 390 395 400
Lys Lys Gly Leu Leu Glu His Gly Arg Asn Trp Ser Ala Ile Ala Arg
405 410 415
Met Val Gly Ser Lys Thr Val Ser Gln Cys Lys Asn Phe Tyr Phe Asn
420 425 430
Tyr Lys Lys Arg Gln Asn Leu Asp Glu Ile Leu Gln Gln His Lys Leu
435 440 445
Lys Met Glu Lys Glu Arg Asn Ala Arg Arg Lys Lys Lys Lys Thr Pro
450 455 460
Ala Ala Ala Ser Glu Glu Thr Ala Phe Pro Pro Ala Ala Glu Asp Glu
465 470 475 480
Glu Met Glu Ala Ser Gly Ala Ser Ala Asn Glu Glu Glu Leu Ala Glu
485 490 495
Glu Ala Glu Ala Ser Gln Ala Ser Gly Asn Glu Val Pro Arg Val Gly
500 505 510
Glu Cys Ser Gly Pro Ala Ala Val Asn Asn Ser Ser Asp Thr Glu Ser
515 520 525
Val Pro Ser Pro Arg Ser Glu Ala Thr Lys Asp Thr Gly Pro Lys Pro
530 535 540
Thr Gly Thr Glu Ala Leu Pro Ala Ala Thr Gln Pro Pro Val Pro Pro
545 550 555 560
Pro Glu Glu Pro Ala Val Ala Pro Ala Glu Pro Ser Pro Val Pro Asp
565 570 575
Ala Ser Gly Pro Pro Ser Pro Glu Pro Ser His His Leu Pro His Pro
580 585 590
Arg Leu Leu Trp Thr Arg Met Asn Lys Lys Pro Arg Leu Leu Gln Leu
595 600 605
Pro Arg Gln Arg Met Pro Arg Ser Arg Ser Leu Arg Pro Arg Arg Ser
610 615 620
Met Trp Glu Lys Pro Glu Glu Pro Glu Ala Ser Glu Lys Pro Pro Lys
625 630 635 640
Ser Val Lys Ser Asp His Lys Lys Glu Thr Glu Glu Glu Pro Glu Asp
645 650 655
Lys Ala Lys Gly Thr Glu Ala Ile Glu Thr Val Ser Glu Ala Pro Leu
660 665 670
Lys Val Glu Lys Ala Gly Ser Lys Ala Ala Val Thr Lys Gly Ser Ser
675 680 685
Ser Gly Ala Thr Gln Asp Ser Asp Ser Ser Ala Thr Cys Ser Ala Asp
690 695 700
Glu Val Asp Glu Pro Glu Gly Gly Asp Lys Gly Arg Leu Leu Ser Pro
705 710 715 720
Arg Pro Ser Leu Leu Thr Pro Ala Gly Asp Pro Arg Ala Ser Thr Ser

```
                725                 730                 735
Pro Gln Lys Pro Leu Asp Leu Lys Gln Leu Lys Gln Arg Ala Ala Ala
            740                 745                 750
Ile Pro Pro Ile Val Thr Lys Val His Glu Pro Pro Arg Glu Asp Thr
        755                 760                 765
Val Pro Pro Lys Pro Val Pro Pro Val Pro Pro Pro Thr Gln His Leu
    770                 775                 780
Gln Pro Glu Gly Asp Val Ser Gln Gln Ser Gly Gly Ser Pro Arg Gly
785                 790                 795                 800
Lys Ser Arg Ser Pro Val Pro Pro Ala Glu Lys Glu Ala Glu Lys Pro
                805                 810                 815
Ala Phe Phe Pro Ala Phe Pro Thr Glu Gly Pro Lys Leu Pro Thr Glu
            820                 825                 830
Pro Pro Arg Trp Ser Ser Gly Leu Pro Phe Pro Ile Pro Pro Arg Glu
        835                 840                 845
Val Ile Lys Thr Ser Pro His Ala Ala Asp Pro Ser Ala Phe Ser Tyr
    850                 855                 860
Thr Pro Pro Gly His Pro Leu Pro Leu Gly Leu His Asp Ser Ala Arg
865                 870                 875                 880
Pro Val Leu Pro Arg Pro Pro Ile Ser Asn Pro Pro Pro Leu Ile Ser
                885                 890                 895
Ser Ala Lys His Pro Gly Val Leu Glu Arg Gln Leu Gly Ala Ile Ser
            900                 905                 910
Gln Gln Gly Met Ser Val Gln Leu Arg Val Pro His Ser Glu His Ala
        915                 920                 925
Lys Ala Pro Met Gly Pro Leu Thr Met Gly Leu Pro Leu Ala Val Asp
    930                 935                 940
Pro Lys Lys Leu Gly Thr Ala Leu Gly Ser Ala Thr Ser Gly Ser Ile
945                 950                 955                 960
Thr Lys Gly Leu Pro Ser Thr Arg Ala Ala Asp Gly Pro Ser Tyr Arg
                965                 970                 975
Gly Ser Ile Thr His Gly Thr Pro Ala Asp Val Leu Tyr Lys Gly Thr
            980                 985                 990
Ile Ser Arg Ile Val Gly Glu Asp Ser Pro Ser Arg Leu Asp Arg Ala
        995                 1000                1005
Arg Glu Asp Thr Leu Pro Lys Gly His Val Ile Tyr Glu Gly Lys Lys
    1010                1015                1020
Gly His Val Leu Ser Tyr Glu Gly Gly Met Ser Val Ser Gln Cys Ser
1025                1030                1035                1040
Lys Glu Asp Gly Arg Ser Ser Ser Gly Pro Pro His Glu Thr Ala Ala
                1045                1050                1055
Pro Lys Arg Thr Tyr Asp Met Met Glu Gly Arg Val Gly Arg Thr Val
            1060                1065                1070
Thr Ser Ala Ser Ile Glu Gly Leu Met Gly Arg Ala Ile Pro Glu Gln
        1075                1080                1085
His Ser Pro His Leu Lys Glu Gln His His Ile Arg Gly Ser Ile Thr
    1090                1095                1100
Gln Gly Ile Pro Arg Ser Tyr Val Glu Ala Gln Glu Asp Tyr Leu Arg
1105                1110                1115                1120
Arg Glu Ala Lys Leu Leu Lys Arg Glu Gly Thr Pro Pro Pro Pro Pro
                1125                1130                1135
Pro Pro Arg Asp Leu Thr Glu Thr Tyr Lys Pro Arg Pro Leu Asp Pro
            1140                1145                1150
Leu Gly Pro Leu Lys Leu Lys Pro Thr His Glu Gly Val Val Ala Thr
        1155                1160                1165
Val Lys Glu Ala Gly Arg Ser Ile His Glu Ile Pro Arg Glu Glu Leu
    1170                1175                1180
Arg Arg Thr Pro Glu Leu Pro Leu Ala Pro Arg Pro Leu Lys Glu Gly
1185                1190                1195                1200
Ser Ile Thr Gln Gly Thr Pro Leu Lys Tyr Asp Ser Gly Ala Pro Ser
```

1205 1210 1215
Thr Gly Thr Lys Lys His Asp Val Arg Ser Ile Ile Gly Ser Pro Gly
1220 1225 1230
Arg Pro Phe Pro Ala Leu His Pro Leu Asp Ile Met Ala Asp Ala Arg
1235 1240 1245
Ala Leu Glu Arg Ala Cys Tyr Glu Glu Ser Leu Lys Ser Arg Ser Gly
1250 1255 1260
Thr Ser Ser Gly Ala Gly Gly Ser Ile Thr Arg Gly Ala Pro Val Val
1265 1270 1275 1280
Val Pro Glu Leu Gly Lys Pro Arg Gln Ser Pro Leu Thr Tyr Glu Asp
1285 1290 1295
His Gly Ala Pro Phe Thr Ser His Leu Pro Arg Gly Ser Pro Val Thr
1300 1305 1310
Thr Arg Glu Pro Thr Pro Arg Leu Gln Glu Gly Ser Leu Leu Ser Ser
1315 1320 1325
Lys Ala Ser Gln Asp Arg Lys Leu Thr Ser Thr Pro Arg Glu Ile Ala
1330 1335 1340
Lys Ser Pro His Ser Thr Val Pro Glu His His Pro His Pro Ile Ser
1345 1350 1355 1360
Pro Tyr Glu His Leu Leu Arg Gly Val Thr Gly Val Asp Leu Tyr Arg
1365 1370 1375
Gly His Ile Pro Leu Ala Phe Asp Pro Thr Ser Ile Pro Arg Gly Ile
1380 1385 1390
Pro Leu Glu Ala Ala Ala Ala Ala Tyr Tyr Leu Pro Arg His Leu Ala
1395 1400 1405
Pro Ser Pro Thr Tyr Pro His Leu Tyr Pro Pro Tyr Leu Ile Arg Gly
1410 1415 1420
Tyr Pro Asp Thr Ala Ala Leu Glu Asn Arg Gln Thr Ile Ile Asn Asp
1425 1430 1435 1440
Tyr Ile Thr Ser Gln Gln Met His His Asn Ala Ala Ser Ala Met Ala
1445 1450 1455
Gln Arg Ala Asp Met Leu Arg Gly Leu Ser Pro Arg Glu Ser Ser Leu
1460 1465 1470
Ala Leu Asn Tyr Ala Ala Gly Pro Arg Gly Ile Ile Asp Leu Ser Gln
1475 1480 1485
Val Pro His Leu Pro Val Leu Val Pro Pro Thr Pro Gly Thr Pro Ala
1490 1495 1500
Thr Ala Ile Asp Arg Leu Ala Tyr Leu Pro Thr Ala Pro Pro Pro Phe
1505 1510 1515 1520
Ser Ser Arg His Ser Ser Ser Pro Leu Ser Pro Gly Gly Pro Thr His
1525 1530 1535
Leu Ala Lys Pro Thr Ala Thr Ser Ser Ser Glu Arg Glu Arg Glu Arg
1540 1545 1550
Glu Arg Glu Arg Asp Lys Ser Ile Leu Thr Ser Thr Thr Thr Val Glu
1555 1560 1565
His Ala Pro Ile Trp Arg Pro Gly Thr Glu Gln Ser Ser Gly Ala Gly
1570 1575 1580
Gly Ser Ser Arg Pro Ala Ser His Thr His Gln His Ser Pro Ile Ser
1585 1590 1595 1600
Pro Arg Thr Gln Asp Ala Leu Gln Gln Arg Pro Ser Val Leu His Asn
1605 1610 1615
Thr Ser Met Lys Gly Val Val Thr Ser Val Glu Pro Gly Thr Pro Thr
1620 1625 1630
Val Leu Arg Trp Ala Arg Ser Thr Ser Thr Ser Ser Pro Val Arg Pro
1635 1640 1645
Ala Ala Thr Phe Pro Pro Ala Thr His Cys Pro Leu Gly Gly Thr Leu
1650 1655 1660
Glu Gly Val Tyr Pro Thr Leu Met Glu Pro Val Leu Leu Pro Lys Glu
1665 1670 1675 1680
Thr Ser Arg Val Ala Arg Pro Glu Arg Ala Arg Val Asp Ala Gly His

1685 1690 1695
Ala Phe Leu Thr Lys Pro Pro Gly Arg Glu Pro Ala Ser Ser Pro Ser
1700 1705 1710
Lys Ser Ser Glu Pro Arg Ser Leu Ala Pro Pro Ser Ser Ser His Thr
1715 1720 1725
Ala Ile Ala Arg Thr Pro Ala Lys Asn Leu Ala Pro His His Ala Ser
1730 1735 1740
Pro Asp Pro Pro Ala Pro Thr Ser Ala Ser Asp Leu His Arg Glu Lys
1745 1750 1755 1760
Thr Gln Ser Lys Pro Phe Ser Ile Gln Glu Leu Glu Leu Arg Ser Leu
1765 1770 1775
Gly Tyr His Ser Gly Ala Gly Tyr Ser Pro Asp Gly Val Glu Pro Ile
1780 1785 1790
Ser Pro Val Ser Ser Pro Ser Leu Thr His Asp Lys Gly Leu Ser Lys
1795 1800 1805
Pro Leu Glu Glu Leu Glu Lys Ser His Leu Glu Gly Glu Leu Arg His
1810 1815 1820
Lys Gln Pro Gly Pro Met Lys Leu Ser Ala Glu Ala Ala His Leu Pro
1825 1830 1835 1840
His Leu Arg Pro Leu Pro Glu Ser Gln Pro Ser Ser Ser Pro Leu Leu
1845 1850 1855
Gln Thr Ala Pro Gly Ile Lys Gly His Gln Arg Val Val Thr Leu Ala
1860 1865 1870
Gln His Ile Ser Glu Val Ile Thr Gln Asp Tyr Thr Arg His His Pro
1875 1880 1885
Gln Gln Leu Ser Gly Pro Leu Pro Ala Pro Leu Tyr Ser Phe Pro Gly
1890 1895 1900
Ala Ser Cys Pro Val Leu Asp Leu Arg Arg Pro Pro Ser Asp Leu Tyr
1905 1910 1915 1920
Leu Pro Pro Pro Asp His Gly Thr Pro Ala Arg Gly Ser Pro His Ser
1925 1930 1935
Glu Gly Gly Lys Arg Ser Pro Glu Pro Ser Lys Thr Ser Val Leu Gly
1940 1945 1950
Ser Ser Glu Asp Ala Ile Glu Pro Val Ser Pro Pro Glu Gly Met Thr
1955 1960 1965
Glu Pro Gly His Ala Arg Ser Thr Ala Tyr Pro Leu Leu Tyr Arg Asp
1970 1975 1980
Gly Glu Gln Gly Glu Pro Arg Met Gly Leu Glu Ser Pro Gly Asn Thr
1985 1990 1995 2000
Ser Gln Pro Pro Thr Phe Phe Ser Lys Leu Thr Glu Ser Asn Ser Ala
2005 2010 2015
Met Val Lys Ser Lys Lys Gln Glu Ile Asn Lys Lys Leu Asn Thr His
2020 2025 2030
Asn Arg Asn Glu Pro Glu Tyr Asn Ile Gly Gln Pro Gly Thr Glu Ile
2035 2040 2045
Phe Asn Met Pro Ala Ile Thr Gly Ala Gly Leu Met Thr Cys Arg Ser
2050 2055 2060
Gln Ala Val Gln Glu His Ala Ser Thr Asn Met Gly Leu Glu Ala Ile
2065 2070 2075 2080
Ile Arg Lys Ala Leu Met Gly Lys Tyr Asp Gln Trp Glu Glu Pro Pro
2085 2090 2095
Pro Leu Gly Ala Asn Ala Phe Asn Pro Leu Asn Ala Ser Ala Ser Leu
2100 2105 2110
Pro Ala Ala Ala Met Pro Ile Thr Thr Ala Asp Gly Arg Ser Asp His
2115 2120 2125
Ala Leu Thr Ser Pro Gly Gly Gly Gly Lys Ala Lys Val Ser Gly Arg
2130 2135 2140
Pro Ser Ser Arg Lys Ala Lys Ser Pro Ala Pro Gly Leu Ala Ser Gly
2145 2150 2155 2160
Asp Arg Pro Pro Ser Val Ser Ser Val His Ser Glu Gly Asp Cys Asn

```
                2165                      2170                      2175
Arg Arg Thr Pro Leu Thr Asn Arg Val Trp Glu Asp Arg Pro Ser Ser
            2180                      2185                      2190
Ala Gly Ser Thr Pro Phe Pro Tyr Asn Pro Leu Ile Met Arg Leu Gln
        2195                      2200                      2205
Ala Gly Val Met Ala Ser Pro Pro Pro Pro Gly Leu Ala Ala Gly Ser
    2210                      2215                      2220
Gly Pro Leu Ala Gly Pro His His Ala Trp Asp Glu Glu Pro Lys Pro
2225                      2230                      2235                      2240
Leu Leu Cys Ser Gln Tyr Glu Thr Leu Ser Asp Ser Glu
                2245                      2250

<210> 10
<211> 7940
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (241)...(7563)

<400> 10
ccaagatggc ggccaaggtg gcgaagcagc agccgcggcg gcggcggcgg ctggagtgag      60
cgtccgactc gccgcgccga acgaggtccc ggtgtagggc cgcgcgccgt ggccgcgtcc     120
cactcctcag gccggggcgc acgtcggctc ccacgcttag ccagctcccg gtggtttcct     180
agaaacatga ttgtttattg gcattgatct cacagtctgg tgaggacttc tttactgata     240
atg tca agt tca ggt tat cct ccc aac caa gga gca ttc agc aca gaa       288
Met Ser Ser Ser Gly Tyr Pro Pro Asn Gln Gly Ala Phe Ser Thr Glu
 1               5                  10                  15

caa agt cgt tat cct cct cac tct gtc cag tat aca ttt ccc aac acc       336
Gln Ser Arg Tyr Pro Pro His Ser Val Gln Tyr Thr Phe Pro Asn Thr
            20                  25                  30

cgc cac cag cag gag ttc gca gtc cct gat tat cgt tcc tct cat ctt       384
Arg His Gln Gln Glu Phe Ala Val Pro Asp Tyr Arg Ser Ser His Leu
        35                  40                  45

gaa gtg agt cag gca tca cag ctt ttg cag caa cag cag cag caa cag       432
Glu Val Ser Gln Ala Ser Gln Leu Leu Gln Gln Gln Gln Gln Gln Gln
    50                  55                  60

ctt cga agg cga cct tcc ttg ctt tca gaa ttt cac cca ggt tct gac       480
Leu Arg Arg Arg Pro Ser Leu Leu Ser Glu Phe His Pro Gly Ser Asp
65                  70                  75                  80

agg cct caa gaa agg aga act agt tat gaa ccg ttt cat cca ggc cca       528
Arg Pro Gln Glu Arg Arg Thr Ser Tyr Glu Pro Phe His Pro Gly Pro
                85                  90                  95

tcc cca gtg gat cat gat tca ctg gaa tcg aag cga cca cgt ctg gaa       576
Ser Pro Val Asp His Asp Ser Leu Glu Ser Lys Arg Pro Arg Leu Glu
            100                 105                 110

cag gtt tct gat tct cat ttt cag cgt gtc agt gct gcg gtt ttg cct       624
Gln Val Ser Asp Ser His Phe Gln Arg Val Ser Ala Ala Val Leu Pro
        115                 120                 125

tta gtg cac ccg ctg cca gaa ggg ctg agg gct tct gca gat gct aag       672
Leu Val His Pro Leu Pro Glu Gly Leu Arg Ala Ser Ala Asp Ala Lys
```

```
    130                 135                 140

aag gat cca gca ttc gga ggc aaa cat gaa gct cca tcc tct cca att      720
Lys Asp Pro Ala Phe Gly Gly Lys His Glu Ala Pro Ser Ser Pro Ile
145                 150                 155                 160

tcg ggg caa cca tgt gga gat gat caa aat gct tca cct tca aaa ctc      768
Ser Gly Gln Pro Cys Gly Asp Asp Gln Asn Ala Ser Pro Ser Lys Leu
                165                 170                 175

tca aag gaa gag tta ata cag agt atg gat cgt gta gat cga gaa att      816
Ser Lys Glu Glu Leu Ile Gln Ser Met Asp Arg Val Asp Arg Glu Ile
            180                 185                 190

gca aaa gta gaa cag cag atc ctt aaa ctg aaa aag aaa caa caa cag      864
Ala Lys Val Glu Gln Gln Ile Leu Lys Leu Lys Lys Lys Gln Gln Gln
        195                 200                 205

ctt gaa gaa gag gca gct aaa cct cct gag cct gag aag ccc gtg tcc      912
Leu Glu Glu Glu Ala Ala Lys Pro Pro Glu Pro Glu Lys Pro Val Ser
    210                 215                 220

cct cct cct gtg gag cag aaa cac cgc agt att gtc caa att att tat      960
Pro Pro Pro Val Glu Gln Lys His Arg Ser Ile Val Gln Ile Ile Tyr
225                 230                 235                 240

gat gag aat cgg aaa aaa gca gaa gaa gct cat aaa att ttt gaa ggt     1008
Asp Glu Asn Arg Lys Lys Ala Glu Glu Ala His Lys Ile Phe Glu Gly
                245                 250                 255

ctt ggc cca aaa gtt gaa ctg cca ctg tat aac cag cca tca gat acc     1056
Leu Gly Pro Lys Val Glu Leu Pro Leu Tyr Asn Gln Pro Ser Asp Thr
            260                 265                 270

aag gtg tac cat gag aac atc aag aca aac cag gtg atg agg aaa aaa     1104
Lys Val Tyr His Glu Asn Ile Lys Thr Asn Gln Val Met Arg Lys Lys
        275                 280                 285

ctc att tta ttt ttt aaa aga aga aat cat gca aga aaa caa agg gaa     1152
Leu Ile Leu Phe Phe Lys Arg Arg Asn His Ala Arg Lys Gln Arg Glu
    290                 295                 300

caa aaa atc tgc cag cgt tat gat cag ctc atg gag gca tgg gag aaa     1200
Gln Lys Ile Cys Gln Arg Tyr Asp Gln Leu Met Glu Ala Trp Glu Lys
305                 310                 315                 320

aaa gtg gac aga ata gaa aat aat cct cgg agg aaa gct aaa gaa agc     1248
Lys Val Asp Arg Ile Glu Asn Asn Pro Arg Arg Lys Ala Lys Glu Ser
                325                 330                 335

aaa aca agg gaa tac tat gaa aag cag ttt cca gaa att cga aaa caa     1296
Lys Thr Arg Glu Tyr Tyr Glu Lys Gln Phe Pro Glu Ile Arg Lys Gln
            340                 345                 350

aga gaa cag caa gaa aga ttt cag cga gtt ggg cag agg gga gct ggt     1344
Arg Glu Gln Gln Glu Arg Phe Gln Arg Val Gly Gln Arg Gly Ala Gly
        355                 360                 365

ctt tca gcc acc att gct agg agt gag cat gag att tct gaa att att     1392
Leu Ser Ala Thr Ile Ala Arg Ser Glu His Glu Ile Ser Glu Ile Ile
```

370 375 380

gat ggg ctc tct gag cag gag aat aat gag aaa caa atg cgg cag ctc 1440
Asp Gly Leu Ser Glu Gln Glu Asn Asn Glu Lys Gln Met Arg Gln Leu
385 390 395 400

tct gtg att cca cct atg atg ttt gat gca gaa caa aga cga gtc aag 1488
Ser Val Ile Pro Pro Met Met Phe Asp Ala Glu Gln Arg Arg Val Lys
405 410 415

ttc att aac atg aat ggg ctt atg gag gac cct atg aaa gtg tat aaa 1536
Phe Ile Asn Met Asn Gly Leu Met Glu Asp Pro Met Lys Val Tyr Lys
420 425 430

gat agg cag ttt atg aat gtt tgg act gac cat gaa aag gag atc ttt 1584
Asp Arg Gln Phe Met Asn Val Trp Thr Asp His Glu Lys Glu Ile Phe
435 440 445

aag gac aag ttt atc cag cat cca aaa aac ttt gga cta att gca tca 1632
Lys Asp Lys Phe Ile Gln His Pro Lys Asn Phe Gly Leu Ile Ala Ser
450 455 460

tac ttg gag agg aag agt gtt cct gat tgt gtt ttg tat tac tat tta 1680
Tyr Leu Glu Arg Lys Ser Val Pro Asp Cys Val Leu Tyr Tyr Tyr Leu
465 470 475 480

acc aag aaa aat gag aat tat aaa gcc ctc gtc aga agg aat tat ggg 1728
Thr Lys Lys Asn Glu Asn Tyr Lys Ala Leu Val Arg Arg Asn Tyr Gly
485 490 495

aaa cgc aga ggc aga aac cag caa att gct cga ccc tcg caa gaa gaa 1776
Lys Arg Arg Gly Arg Asn Gln Gln Ile Ala Arg Pro Ser Gln Glu Glu
500 505 510

aaa gta gaa gaa aaa gaa gag gat aaa gca gaa aaa aca gaa aaa aaa 1824
Lys Val Glu Glu Lys Glu Glu Asp Lys Ala Glu Lys Thr Glu Lys Lys
515 520 525

gaa gaa gaa aag aaa gat gaa gag gaa aaa gat gaa aaa gaa gac tcc 1872
Glu Glu Glu Lys Lys Asp Glu Glu Glu Lys Asp Glu Lys Glu Asp Ser
530 535 540

aaa gaa aat acc aag gaa aag gac aag ata gat ggt aca gca gaa gaa 1920
Lys Glu Asn Thr Lys Glu Lys Asp Lys Ile Asp Gly Thr Ala Glu Glu
545 550 555 560

act gag gaa aga gag caa gcc aca ccc cgg ggg cga aag act gcc aac 1968
Thr Glu Glu Arg Glu Gln Ala Thr Pro Arg Gly Arg Lys Thr Ala Asn
565 570 575

agt cag ggc cgc cgt aag ggc cgg atc acc agg tcc atg aca aac gaa 2016
Ser Gln Gly Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Thr Asn Glu
580 585 590

gct gca gct gcc agt gct gca gcc gca gcg gct act gaa gag ccc cca 2064
Ala Ala Ala Ala Ser Ala Ala Ala Ala Ala Ala Thr Glu Glu Pro Pro
595 600 605

cca cct ctg cca ccg cca cca gaa ccc att tct aca gag cct gtg gag 2112
Pro Pro Leu Pro Pro Pro Pro Glu Pro Ile Ser Thr Glu Pro Val Glu

```
    610                 615                 620

acc tct cga tgg aca gaa gaa gaa atg gaa gtt gct aaa aaa ggt cta      2160
Thr Ser Arg Trp Thr Glu Glu Glu Met Glu Val Ala Lys Lys Gly Leu
625                 630                 635                 640

gta gaa cat ggt cgt aac tgg gca gca att gct aaa atg gtg gga acg      2208
Val Glu His Gly Arg Asn Trp Ala Ala Ile Ala Lys Met Val Gly Thr
                645                 650                 655

aaa agt gaa gct caa tgt aaa aac ttc tat ttt aac tat aaa agg cga      2256
Lys Ser Glu Ala Gln Cys Lys Asn Phe Tyr Phe Asn Tyr Lys Arg Arg
            660                 665                 670

cac aat ctt gac aac ctc tta cag cag cat aaa cag aaa act tca cga      2304
His Asn Leu Asp Asn Leu Leu Gln Gln His Lys Gln Lys Thr Ser Arg
        675                 680                 685

aaa cct cgt gaa gag cga gat gtg tct caa tgt gaa agt gtc gct tcc      2352
Lys Pro Arg Glu Glu Arg Asp Val Ser Gln Cys Glu Ser Val Ala Ser
    690                 695                 700

act gtt tct gct cag gag gat gaa gat att gaa gcc tcc aat gaa gaa      2400
Thr Val Ser Ala Gln Glu Asp Glu Asp Ile Glu Ala Ser Asn Glu Glu
705                 710                 715                 720

gaa aat cca gaa gac agc gaa gtt gaa gct gtc aag ccc agc gag gac      2448
Glu Asn Pro Glu Asp Ser Glu Val Glu Ala Val Lys Pro Ser Glu Asp
                725                 730                 735

agt cct gaa aat gct act tct cga gga aac aca gaa cct gcg gtt gag      2496
Ser Pro Glu Asn Ala Thr Ser Arg Gly Asn Thr Glu Pro Ala Val Glu
            740                 745                 750

ctt gag ccc acc acg gaa act gca ccc agt aca tct ccc tcc tta gca      2544
Leu Glu Pro Thr Thr Glu Thr Ala Pro Ser Thr Ser Pro Ser Leu Ala
        755                 760                 765

gtt cca agt aca aaa cca gct gaa gat gaa agt gtg gag acc cag gtg      2592
Val Pro Ser Thr Lys Pro Ala Glu Asp Glu Ser Val Glu Thr Gln Val
    770                 775                 780

aat gac agc atc agt gct gag aca gca gag cag atg gat gta gat cag      2640
Asn Asp Ser Ile Ser Ala Glu Thr Ala Glu Gln Met Asp Val Asp Gln
785                 790                 795                 800

cag gag cac agt gct gaa gag ggt tct gtt tgt gat ccc cca ccc gct      2688
Gln Glu His Ser Ala Glu Glu Gly Ser Val Cys Asp Pro Pro Pro Ala
                805                 810                 815

acc aaa gct gac tct gtg gac gtt gaa gtg agg gtg cca gaa aac cat      2736
Thr Lys Ala Asp Ser Val Asp Val Glu Val Arg Val Pro Glu Asn His
            820                 825                 830

gca tct aaa gtt gaa ggt gat aat acc aaa gaa aga gac ttg gat aga      2784
Ala Ser Lys Val Glu Gly Asp Asn Thr Lys Glu Arg Asp Leu Asp Arg
        835                 840                 845

gcc agt gag aag gtg gaa cct aga gat gaa gat ttg gtg gta gct cag      2832
Ala Ser Glu Lys Val Glu Pro Arg Asp Glu Asp Leu Val Val Ala Gln
```

```
    850                 855                 860

caa ata aat gcc caa agg ccc gag ccc cag tca gac aat gat tcc agt      2880
Gln Ile Asn Ala Gln Arg Pro Glu Pro Gln Ser Asp Asn Asp Ser Ser
865                 870                 875                 880

gcc acg tgc agc gct gat gag gat gtg gat gga gag cca gag agg cag      2928
Ala Thr Cys Ser Ala Asp Glu Asp Val Asp Gly Glu Pro Glu Arg Gln
                885                 890                 895

aga atg ttt cct atg gac tca aag cct tca ctg tta aac ccc act gga      2976
Arg Met Phe Pro Met Asp Ser Lys Pro Ser Leu Leu Asn Pro Thr Gly
            900                 905                 910

tct ata ctc gtc tca tct ccg tta aaa cca aat cca ctg gat ctg cca      3024
Ser Ile Leu Val Ser Ser Pro Leu Lys Pro Asn Pro Leu Asp Leu Pro
        915                 920                 925

cag ctt cag cat cga gct gct gtt atc cca cca atg gta tcc tgc acc      3072
Gln Leu Gln His Arg Ala Ala Val Ile Pro Pro Met Val Ser Cys Thr
    930                 935                 940

cca tgt aac ata cca att gga acc cca gtg agc ggc tat gct ctc tac      3120
Pro Cys Asn Ile Pro Ile Gly Thr Pro Val Ser Gly Tyr Ala Leu Tyr
945                 950                 955                 960

cag cga cac att aaa gca atg cat gag tca gca ctc ctg gag gag cag      3168
Gln Arg His Ile Lys Ala Met His Glu Ser Ala Leu Leu Glu Glu Gln
                965                 970                 975

cgg cag aga caa gaa cag ata gat ttg gaa tgt aga agt tct aca agt      3216
Arg Gln Arg Gln Glu Gln Ile Asp Leu Glu Cys Arg Ser Ser Thr Ser
            980                 985                 990

cca tgt ggc aca tcc aag agt cca aac aga gag tgg gaa gtc ctt cag      3264
Pro Cys Gly Thr Ser Lys Ser Pro Asn Arg Glu Trp Glu Val Leu Gln
        995                 1000                1005

cct gct cca cat caa ttg ata act aat ctc cct gaa ggc gtt cgg ctt      3312
Pro Ala Pro His Gln Leu Ile Thr Asn Leu Pro Glu Gly Val Arg Leu
    1010                1015                1020

ccg aca act cga cca acc agg cca ccg ccc cct ctc atc ccg tca tcc      3360
Pro Thr Thr Arg Pro Thr Arg Pro Pro Pro Pro Leu Ile Pro Ser Ser
1025                1030                1035                1040

aaa acc aca gtg gct tca gaa aaa cca tct ttt ata atg gga ggc tcc      3408
Lys Thr Thr Val Ala Ser Glu Lys Pro Ser Phe Ile Met Gly Gly Ser
                1045                1050                1055

atc tca cag gga aca cca ggc act tat ttg act tct cat aat cag gct      3456
Ile Ser Gln Gly Thr Pro Gly Thr Tyr Leu Thr Ser His Asn Gln Ala
            1060                1065                1070

tcc tac act caa gaa aca ccc aag ccg tca gta gga tct atc tct ctt      3504
Ser Tyr Thr Gln Glu Thr Pro Lys Pro Ser Val Gly Ser Ile Ser Leu
        1075                1080                1085

gga ctg cca cgg caa cag gaa tct gcc aaa tca gct act ttg ccc tac      3552
Gly Leu Pro Arg Gln Gln Glu Ser Ala Lys Ser Ala Thr Leu Pro Tyr
```

1090 1095 1100

atc aag cag gaa gaa ttt tct ccc cga agc caa aac tca caa cct gag 3600
Ile Lys Gln Glu Glu Phe Ser Pro Arg Ser Gln Asn Ser Gln Pro Glu
1105 1110 1115 1120

ggt ctg ttg gtc agg gcc caa cat gaa ggt gta gtc aga ggt acc gca 3648
Gly Leu Leu Val Arg Ala Gln His Glu Gly Val Val Arg Gly Thr Ala
1125 1130 1135

gga gcc ata caa gaa gga agt ata act cgg gga act cca acc agc aaa 3696
Gly Ala Ile Gln Glu Gly Ser Ile Thr Arg Gly Thr Pro Thr Ser Lys
1140 1145 1150

att tca gtg gag agc att cca tcc cta cgg ggc tct atc act cag ggc 3744
Ile Ser Val Glu Ser Ile Pro Ser Leu Arg Gly Ser Ile Thr Gln Gly
1155 1160 1165

acc ccg gct ctg ccc cag act ggc ata cca aca gag gct ttg gtg aag 3792
Thr Pro Ala Leu Pro Gln Thr Gly Ile Pro Thr Glu Ala Leu Val Lys
1170 1175 1180

ggg tcc att tcg aga atg ccc att gaa gac agc agt cct gag aaa ggc 3840
Gly Ser Ile Ser Arg Met Pro Ile Glu Asp Ser Ser Pro Glu Lys Gly
1185 1190 1195 1200

aga gag gaa gct gca tcc aaa ggc cat gtt att tat gaa ggc aaa agt 3888
Arg Glu Glu Ala Ala Ser Lys Gly His Val Ile Tyr Glu Gly Lys Ser
1205 1210 1215

gga cat atc ttg tca tat gat aat att aag aat gcc cga gaa ggg act 3936
Gly His Ile Leu Ser Tyr Asp Asn Ile Lys Asn Ala Arg Glu Gly Thr
1220 1225 1230

agg agt cca aga aca gct cat gaa atc agt tta aag aga agc tat gaa 3984
Arg Ser Pro Arg Thr Ala His Glu Ile Ser Leu Lys Arg Ser Tyr Glu
1235 1240 1245

tca gtg gaa gga aat ata aag caa ggg atg tca atg agg gag tct cct 4032
Ser Val Glu Gly Asn Ile Lys Gln Gly Met Ser Met Arg Glu Ser Pro
1250 1255 1260

gta tca gca ccg tta gag ggg ctg ata tgc cga gca tta ccc agg ggg 4080
Val Ser Ala Pro Leu Glu Gly Leu Ile Cys Arg Ala Leu Pro Arg Gly
1265 1270 1275 1280

agt cct cat tct gac ctc aaa gaa agg act gta ttg tct ggc tcc ata 4128
Ser Pro His Ser Asp Leu Lys Glu Arg Thr Val Leu Ser Gly Ser Ile
1285 1290 1295

atg cag ggg aca cca aga gca aca act gaa agc ttt gaa gat ggc ctt 4176
Met Gln Gly Thr Pro Arg Ala Thr Thr Glu Ser Phe Glu Asp Gly Leu
1300 1305 1310

aaa tat ccc aaa caa att aaa agg gaa agt cct ccc ata cga gca ttt 4224
Lys Tyr Pro Lys Gln Ile Lys Arg Glu Ser Pro Pro Ile Arg Ala Phe
1315 1320 1325

gaa ggt gcc att acc aaa gga aaa cca tat gat ggc atc acc acc atc 4272
Glu Gly Ala Ile Thr Lys Gly Lys Pro Tyr Asp Gly Ile Thr Thr Ile

1330 1335 1340

aaa gaa atg ggg cgt tcc att cat gag att cca agg caa gat att tta 4320
Lys Glu Met Gly Arg Ser Ile His Glu Ile Pro Arg Gln Asp Ile Leu
1345 1350 1355 1360

act cag gaa agt cgg aaa act cca gaa gtg gtc cag agc aca cgg ccg 4368
Thr Gln Glu Ser Arg Lys Thr Pro Glu Val Val Gln Ser Thr Arg Pro
1365 1370 1375

ata att gag ggt tcc att tcc cag ggc aca cca ata aag ttt gac aac 4416
Ile Ile Glu Gly Ser Ile Ser Gln Gly Thr Pro Ile Lys Phe Asp Asn
1380 1385 1390

aac tca ggt caa tct gcc atc aaa cac aat gtc aaa tcc tta atc acg 4464
Asn Ser Gly Gln Ser Ala Ile Lys His Asn Val Lys Ser Leu Ile Thr
1395 1400 1405

ggg cct agc aaa cta tcc cgt gga atg cct ccg ctg gaa att gtg cca 4512
Gly Pro Ser Lys Leu Ser Arg Gly Met Pro Pro Leu Glu Ile Val Pro
1410 1415 1420

gag aac ata aaa gtg gta gaa cgg gga aaa tat gag gat gtg aaa gca 4560
Glu Asn Ile Lys Val Val Glu Arg Gly Lys Tyr Glu Asp Val Lys Ala
1425 1430 1435 1440

ggc gag acc gtg cgt tcc cgg cac acg tca gtg gta agc tct ggc ccc 4608
Gly Glu Thr Val Arg Ser Arg His Thr Ser Val Val Ser Ser Gly Pro
1445 1450 1455

tcc gtt ctt agg tcc aca ctg cat gaa gct ccc aaa gca caa ctg agc 4656
Ser Val Leu Arg Ser Thr Leu His Glu Ala Pro Lys Ala Gln Leu Ser
1460 1465 1470

cct ggg att tat gat gac acc agt gca cgg agg acc cct gtg agt tat 4704
Pro Gly Ile Tyr Asp Asp Thr Ser Ala Arg Arg Thr Pro Val Ser Tyr
1475 1480 1485

caa aac acc atg tcc aga ggc tca ccc atg atg aac aga act tct gat 4752
Gln Asn Thr Met Ser Arg Gly Ser Pro Met Met Asn Arg Thr Ser Asp
1490 1495 1500

gtt aca att cct cct aac aag tct acc aat cat gaa agg aaa tcg aca 4800
Val Thr Ile Pro Pro Asn Lys Ser Thr Asn His Glu Arg Lys Ser Thr
1505 1510 1515 1520

ctg acc cct acc cag agg gaa agt atc cca gcg aag tct cca gtg cct 4848
Leu Thr Pro Thr Gln Arg Glu Ser Ile Pro Ala Lys Ser Pro Val Pro
1525 1530 1535

ggg gtg gac cct gtc gtg agc cac agt ccg ttt gat ccc cat cac aga 4896
Gly Val Asp Pro Val Val Ser His Ser Pro Phe Asp Pro His His Arg
1540 1545 1550

ggc agc act gca ggc gag gtt tat tgg agc cac ctg ccc acg caa ttg 4944
Gly Ser Thr Ala Gly Glu Val Tyr Trp Ser His Leu Pro Thr Gln Leu
1555 1560 1565

gat cca gcc atg cct ttt cac agg gct ttg gat cct gca gcg gct gct 4992
Asp Pro Ala Met Pro Phe His Arg Ala Leu Asp Pro Ala Ala Ala Ala

1570 1575 1580

tac ctg ttt cag aga cag ctt tca cca act cca ggt tac cca agt cag 5040
Tyr Leu Phe Gln Arg Gln Leu Ser Pro Thr Pro Gly Tyr Pro Ser Gln
1585 1590 1595 1600

tat cag ctt tac gca atg gag aac aca aga cag aca atc tta aat gat 5088
Tyr Gln Leu Tyr Ala Met Glu Asn Thr Arg Gln Thr Ile Leu Asn Asp
1605 1610 1615

tac att acc tca caa cag atg caa gtg aac ttg cgt cca gat gtg gcc 5136
Tyr Ile Thr Ser Gln Gln Met Gln Val Asn Leu Arg Pro Asp Val Ala
1620 1625 1630

aga gga ctc tcc cca aga gag cag cca ctg ggt ctc cca tac cca gca 5184
Arg Gly Leu Ser Pro Arg Glu Gln Pro Leu Gly Leu Pro Tyr Pro Ala
1635 1640 1645

acg aga gga atc att gac ctg acc aat atg cct cca aca att tta gtg 5232
Thr Arg Gly Ile Ile Asp Leu Thr Asn Met Pro Pro Thr Ile Leu Val
1650 1655 1660

cct cat cca ggg gga aca agc act cct ccc atg gac aga atc act tat 5280
Pro His Pro Gly Gly Thr Ser Thr Pro Pro Met Asp Arg Ile Thr Tyr
1665 1670 1675 1680

att cct ggt aca cag att act ttc cct ccc agg ccg tac aac tct gct 5328
Ile Pro Gly Thr Gln Ile Thr Phe Pro Pro Arg Pro Tyr Asn Ser Ala
1685 1690 1695

tcc atg tct cca gga cac cca aca cac ctt gca gct gct gca agt gct 5376
Ser Met Ser Pro Gly His Pro Thr His Leu Ala Ala Ala Ala Ser Ala
1700 1705 1710

gag agg gaa cgg gaa cgg gag cgg gag aag gag cgg gag cgg gaa cgg 5424
Glu Arg Glu Arg Glu Arg Glu Arg Glu Lys Glu Arg Glu Arg Glu Arg
1715 1720 1725

att gct gca gct tcc tcc gac ctc tac ctg cgg cca ggc tca gaa cag 5472
Ile Ala Ala Ala Ser Ser Asp Leu Tyr Leu Arg Pro Gly Ser Glu Gln
1730 1735 1740

cct ggc cga cct ggc agt cat gga tat gtt cgc tcc cct tcc cct tca 5520
Pro Gly Arg Pro Gly Ser His Gly Tyr Val Arg Ser Pro Ser Pro Ser
1745 1750 1755 1760

gta aga act cag gag acc atg ttg caa cag aga ccc agt gtt ttc caa 5568
Val Arg Thr Gln Glu Thr Met Leu Gln Gln Arg Pro Ser Val Phe Gln
1765 1770 1775

gga acc aat gga acc agt gta atc aca cct ttg gat cca act gct cag 5616
Gly Thr Asn Gly Thr Ser Val Ile Thr Pro Leu Asp Pro Thr Ala Gln
1780 1785 1790

cta cga atc atg cca ctg cct gct ggg ggc cct tca ata agc caa ggc 5664
Leu Arg Ile Met Pro Leu Pro Ala Gly Gly Pro Ser Ile Ser Gln Gly
1795 1800 1805

ctg cca gcc tcc cgt tac aac act gct gcg gat gcc ctg gct gct ctt 5712
Leu Pro Ala Ser Arg Tyr Asn Thr Ala Ala Asp Ala Leu Ala Ala Leu

1810 1815 1820

gtg gat gct gca gct tct gca ccc cag atg gat gtg tcc aaa aca aaa 5760
Val Asp Ala Ala Ala Ser Ala Pro Gln Met Asp Val Ser Lys Thr Lys
1825 1830 1835 1840

gag agt aag cat gaa gct gcc agg tta gaa gaa aat ttg aga agc agg 5808
Glu Ser Lys His Glu Ala Ala Arg Leu Glu Glu Asn Leu Arg Ser Arg
1845 1850 1855

tca gca gca gtt agt gaa cag cag cag cta gag cag aaa acc ctg gag 5856
Ser Ala Ala Val Ser Glu Gln Gln Gln Leu Glu Gln Lys Thr Leu Glu
1860 1865 1870

gtg gag aag aga tct gtt cag tgt tta tac act tct tca gcc ttt cca 5904
Val Glu Lys Arg Ser Val Gln Cys Leu Tyr Thr Ser Ser Ala Phe Pro
1875 1880 1885

agt ggc aag ccc cag cct cat tct tca gta gtt tat tct gag gct ggg 5952
Ser Gly Lys Pro Gln Pro His Ser Ser Val Val Tyr Ser Glu Ala Gly
1890 1895 1900

aaa gat aaa ggg cct cct cca aaa tcc aga tat gag gaa gag cta agg 6000
Lys Asp Lys Gly Pro Pro Pro Lys Ser Arg Tyr Glu Glu Glu Leu Arg
1905 1910 1915 1920

acc aga ggg aag act acc att act gca gct aac ttc ata gac gtg atc 6048
Thr Arg Gly Lys Thr Thr Ile Thr Ala Ala Asn Phe Ile Asp Val Ile
1925 1930 1935

atc acc cgg caa att gcc tcg gac aag gat gcg agg gaa cgt ggc tct 6096
Ile Thr Arg Gln Ile Ala Ser Asp Lys Asp Ala Arg Glu Arg Gly Ser
1940 1945 1950

caa agt tca gac tct tct agt agc tta tct tct cac agg tat gaa aca 6144
Gln Ser Ser Asp Ser Ser Ser Ser Leu Ser Ser His Arg Tyr Glu Thr
1955 1960 1965

cct agc gat gct att gag gtg ata agt cct gcc agc tca cct gcg cca 6192
Pro Ser Asp Ala Ile Glu Val Ile Ser Pro Ala Ser Ser Pro Ala Pro
1970 1975 1980

ccc cag gag aaa ctg cag acc tat cag cca gag gtt gtt aag gca aat 6240
Pro Gln Glu Lys Leu Gln Thr Tyr Gln Pro Glu Val Val Lys Ala Asn
1985 1990 1995 2000

caa gcg gaa aat gat cct acc aga caa tat gaa gga cca tta cat cac 6288
Gln Ala Glu Asn Asp Pro Thr Arg Gln Tyr Glu Gly Pro Leu His His
2005 2010 2015

tat cga cca cag cag gaa tca cca tct ccc caa caa cag ctg ccc cct 6336
Tyr Arg Pro Gln Gln Glu Ser Pro Ser Pro Gln Gln Gln Leu Pro Pro
2020 2025 2030

tct tca cag gca gag gga atg ggg caa gtg ccc agg acc cat cgg ctg 6384
Ser Ser Gln Ala Glu Gly Met Gly Gln Val Pro Arg Thr His Arg Leu
2035 2040 2045

atc aca ctt gct gat cac atc tgt caa att atc aca caa gat ttt gct 6432
Ile Thr Leu Ala Asp His Ile Cys Gln Ile Ile Thr Gln Asp Phe Ala

2050 2055 2060

aga aat caa gtt tcc tcg cag act ccc cag cag cct cct act tct aca 6480
Arg Asn Gln Val Ser Ser Gln Thr Pro Gln Gln Pro Pro Thr Ser Thr
2065 2070 2075 2080

ttc cag aac tca cct tct gct ttg gta tct aca cct gtg agg act aaa 6528
Phe Gln Asn Ser Pro Ser Ala Leu Val Ser Thr Pro Val Arg Thr Lys
2085 2090 2095

aca tca aac cgt tac agc cca gaa tcc cag gct cag tct gtc cat cat 6576
Thr Ser Asn Arg Tyr Ser Pro Glu Ser Gln Ala Gln Ser Val His His
2100 2105 2110

caa aga cca ggt tca agg gtc tct cca gaa aat ctt gtg gac aaa tcc 6624
Gln Arg Pro Gly Ser Arg Val Ser Pro Glu Asn Leu Val Asp Lys Ser
2115 2120 2125

agg gga agt agg cct gga aaa tcc cca gag agg agt cac gtc tct tcc 6672
Arg Gly Ser Arg Pro Gly Lys Ser Pro Glu Arg Ser His Val Ser Ser
2130 2135 2140

gag ccc tac gag ccc atc tcc cca ccc cag gtt ccg gtt gtg cat gag 6720
Glu Pro Tyr Glu Pro Ile Ser Pro Pro Gln Val Pro Val Val His Glu
2145 2150 2155 2160

aaa cag gac agc ttg ctg ctc ttg tct cag agg ggc gca gag cct gca 6768
Lys Gln Asp Ser Leu Leu Leu Leu Ser Gln Arg Gly Ala Glu Pro Ala
2165 2170 2175

gag cag agg aat gat gcc cgc tca cca ggg agt ata agc tac ttg cct 6816
Glu Gln Arg Asn Asp Ala Arg Ser Pro Gly Ser Ile Ser Tyr Leu Pro
2180 2185 2190

tca ttc ttc acc aag ctt gaa aat aca tca ccc atg gtt aaa tca aag 6864
Ser Phe Phe Thr Lys Leu Glu Asn Thr Ser Pro Met Val Lys Ser Lys
2195 2200 2205

aag cag gag att ttt cgt aag ttg aac tcc tct ggt gga ggt gac tct 6912
Lys Gln Glu Ile Phe Arg Lys Leu Asn Ser Ser Gly Gly Gly Asp Ser
2210 2215 2220

gat atg gca gct gct cag cca gga act gag atc ttt aat ctg cca gca 6960
Asp Met Ala Ala Ala Gln Pro Gly Thr Glu Ile Phe Asn Leu Pro Ala
2225 2230 2235 2240

gtt act acg tca ggc tca gtt agc tct aga ggc cat tct ttt gct gat 7008
Val Thr Thr Ser Gly Ser Val Ser Ser Arg Gly His Ser Phe Ala Asp
2245 2250 2255

cct gcc agt aat ctt ggg ctg gaa gac att atc agg aag gct ctc atg 7056
Pro Ala Ser Asn Leu Gly Leu Glu Asp Ile Ile Arg Lys Ala Leu Met
2260 2265 2270

gga agc ttt gat gac aaa gtt gag gat cat gga gtt gtc atg tcc cag 7104
Gly Ser Phe Asp Asp Lys Val Glu Asp His Gly Val Val Met Ser Gln
2275 2280 2285

cct atg gga gta gtg cct ggt act gcc aac acc tca gtt gtg acc agt 7152
Pro Met Gly Val Val Pro Gly Thr Ala Asn Thr Ser Val Val Thr Ser

2290 2295 2300

ggt gag aca cga aga gag gaa ggg gac cca tca cct cat tca gga gga 7200
Gly Glu Thr Arg Arg Glu Glu Gly Asp Pro Ser Pro His Ser Gly Gly
2305 2310 2315 2320

gtt tgc aaa cca aag ctg atc agc aag tca aac agc agg aaa tct aag 7248
Val Cys Lys Pro Lys Leu Ile Ser Lys Ser Asn Ser Arg Lys Ser Lys
2325 2330 2335

tct cct ata cct ggg caa ggc tac tta gga acg gaa cgg ccc tct tca 7296
Ser Pro Ile Pro Gly Gln Gly Tyr Leu Gly Thr Glu Arg Pro Ser Ser
2340 2345 2350

gtc tcc tct gta cat tca gaa ggg gat tac cat agg cag acg cca ggg 7344
Val Ser Ser Val His Ser Glu Gly Asp Tyr His Arg Gln Thr Pro Gly
2355 2360 2365

tgg gcc tgg gaa gac agg ccc tct tca aca ggc tca act cag ttt cct 7392
Trp Ala Trp Glu Asp Arg Pro Ser Ser Thr Gly Ser Thr Gln Phe Pro
2370 2375 2380

tat aac cct ctg act atg cgg atg ctc agc agt act cca cca aca ccg 7440
Tyr Asn Pro Leu Thr Met Arg Met Leu Ser Ser Thr Pro Pro Thr Pro
2385 2390 2395 2400

att gca tgt gct ccc tct gcg gtg aac caa gca gct cct cac caa cag 7488
Ile Ala Cys Ala Pro Ser Ala Val Asn Gln Ala Ala Pro His Gln Gln
2405 2410 2415

aac agg atc tgg gag cga gag cct gcc cca ctg ctc tca gca cag tac 7536
Asn Arg Ile Trp Glu Arg Glu Pro Ala Pro Leu Leu Ser Ala Gln Tyr
2420 2425 2430

gag acc ctg tcg gat agt gat gac tga actgcacaaa gtgaggggaa 7583
Glu Thr Leu Ser Asp Ser Asp Asp *
2435 2440

cagggtgcag gagagggatc tctagttttt gtggtttaat ttttagtagc aggtcaaaaa 7643
cctgccctcc tgtgacttat tccctgagac ttttcaggag agccagccca cagatgatga 7703
agaaatgatg gaagttcatt tggagagtca aatgggaaaa aaacaaacaa aaaactgcct 7763
ttgatacagg caattcagtg gactataata atagtggagg gttgagatgt agagttttta 7823
aaaagtgaac agttgctgtt cttacatctg taaagaaaac cataatgtct ttaaatcact 7883
cttctgtaaa tagatgacct ttttgcagtg taaaaaaaaa aaaaaaaaaa aaaaaaa 7940

<210> 11
<211> 2440
<212> PRT
<213> Homo sapiens

<400> 11
Met Ser Ser Ser Gly Tyr Pro Pro Asn Gln Gly Ala Phe Ser Thr Glu
1 5 10 15
Gln Ser Arg Tyr Pro Pro His Ser Val Gln Tyr Thr Phe Pro Asn Thr
20 25 30
Arg His Gln Gln Glu Phe Ala Val Pro Asp Tyr Arg Ser Ser His Leu
35 40 45
Glu Val Ser Gln Ala Ser Gln Leu Leu Gln Gln Gln Gln Gln Gln Gln
50 55 60
Leu Arg Arg Arg Pro Ser Leu Leu Ser Glu Phe His Pro Gly Ser Asp

65 70 75 80
Arg Pro Gln Glu Arg Arg Thr Ser Tyr Glu Pro Phe His Pro Gly Pro
85 90 95
Ser Pro Val Asp His Asp Ser Leu Glu Ser Lys Arg Pro Arg Leu Glu
100 105 110
Gln Val Ser Asp Ser His Phe Gln Arg Val Ser Ala Ala Val Leu Pro
115 120 125
Leu Val His Pro Leu Pro Glu Gly Leu Arg Ala Ser Ala Asp Ala Lys
130 135 140
Lys Asp Pro Ala Phe Gly Gly Lys His Glu Ala Pro Ser Ser Pro Ile
145 150 155 160
Ser Gly Gln Pro Cys Gly Asp Asp Gln Asn Ala Ser Pro Ser Lys Leu
165 170 175
Ser Lys Glu Glu Leu Ile Gln Ser Met Asp Arg Val Asp Arg Glu Ile
180 185 190
Ala Lys Val Glu Gln Gln Ile Leu Lys Leu Lys Lys Lys Gln Gln Gln
195 200 205
Leu Glu Glu Glu Ala Ala Lys Pro Pro Glu Pro Glu Lys Pro Val Ser
210 215 220
Pro Pro Pro Val Glu Gln Lys His Arg Ser Ile Val Gln Ile Ile Tyr
225 230 235 240
Asp Glu Asn Arg Lys Lys Ala Glu Glu Ala His Lys Ile Phe Glu Gly
245 250 255
Leu Gly Pro Lys Val Glu Leu Pro Leu Tyr Asn Gln Pro Ser Asp Thr
260 265 270
Lys Val Tyr His Glu Asn Ile Lys Thr Asn Gln Val Met Arg Lys Lys
275 280 285
Leu Ile Leu Phe Phe Lys Arg Arg Asn His Ala Arg Lys Gln Arg Glu
290 295 300
Gln Lys Ile Cys Gln Arg Tyr Asp Gln Leu Met Glu Ala Trp Glu Lys
305 310 315 320
Lys Val Asp Arg Ile Glu Asn Asn Pro Arg Arg Lys Ala Lys Glu Ser
325 330 335
Lys Thr Arg Glu Tyr Tyr Glu Lys Gln Phe Pro Glu Ile Arg Lys Gln
340 345 350
Arg Glu Gln Gln Glu Arg Phe Gln Arg Val Gly Gln Arg Gly Ala Gly
355 360 365
Leu Ser Ala Thr Ile Ala Arg Ser Glu His Glu Ile Ser Glu Ile Ile
370 375 380
Asp Gly Leu Ser Glu Gln Glu Asn Asn Glu Lys Gln Met Arg Gln Leu
385 390 395 400
Ser Val Ile Pro Pro Met Met Phe Asp Ala Glu Gln Arg Arg Val Lys
405 410 415
Phe Ile Asn Met Asn Gly Leu Met Glu Asp Pro Met Lys Val Tyr Lys
420 425 430
Asp Arg Gln Phe Met Asn Val Trp Thr Asp His Glu Lys Glu Ile Phe
435 440 445
Lys Asp Lys Phe Ile Gln His Pro Lys Asn Phe Gly Leu Ile Ala Ser
450 455 460
Tyr Leu Glu Arg Lys Ser Val Pro Asp Cys Val Leu Tyr Tyr Tyr Leu
465 470 475 480
Thr Lys Lys Asn Glu Asn Tyr Lys Ala Leu Val Arg Arg Asn Tyr Gly
485 490 495
Lys Arg Arg Gly Arg Asn Gln Gln Ile Ala Arg Pro Ser Gln Glu Glu
500 505 510
Lys Val Glu Glu Lys Glu Glu Asp Lys Ala Glu Lys Thr Glu Lys Lys
515 520 525
Glu Glu Glu Lys Lys Asp Glu Glu Glu Lys Asp Glu Lys Glu Asp Ser
530 535 540
Lys Glu Asn Thr Lys Glu Lys Asp Lys Ile Asp Gly Thr Ala Glu Glu

545 550 555 560
Thr Glu Glu Arg Glu Gln Ala Thr Pro Arg Gly Arg Lys Thr Ala Asn
565 570 575
Ser Gln Gly Arg Arg Lys Gly Arg Ile Thr Arg Ser Met Thr Asn Glu
580 585 590
Ala Ala Ala Ala Ser Ala Ala Ala Ala Ala Ala Thr Glu Glu Pro Pro
595 600 605
Pro Pro Leu Pro Pro Pro Pro Glu Pro Ile Ser Thr Glu Pro Val Glu
610 615 620
Thr Ser Arg Trp Thr Glu Glu Glu Met Glu Val Ala Lys Lys Gly Leu
625 630 635 640
Val Glu His Gly Arg Asn Trp Ala Ala Ile Ala Lys Met Val Gly Thr
645 650 655
Lys Ser Glu Ala Gln Cys Lys Asn Phe Tyr Phe Asn Tyr Lys Arg Arg
660 665 670
His Asn Leu Asp Asn Leu Leu Gln Gln His Lys Gln Lys Thr Ser Arg
675 680 685
Lys Pro Arg Glu Glu Arg Asp Val Ser Gln Cys Glu Ser Val Ala Ser
690 695 700
Thr Val Ser Ala Gln Glu Asp Glu Asp Ile Glu Ala Ser Asn Glu Glu
705 710 715 720
Glu Asn Pro Glu Asp Ser Glu Val Glu Ala Val Lys Pro Ser Glu Asp
725 730 735
Ser Pro Glu Asn Ala Thr Ser Arg Gly Asn Thr Glu Pro Ala Val Glu
740 745 750
Leu Glu Pro Thr Thr Glu Thr Ala Pro Ser Thr Ser Pro Ser Leu Ala
755 760 765
Val Pro Ser Thr Lys Pro Ala Glu Asp Glu Ser Val Glu Thr Gln Val
770 775 780
Asn Asp Ser Ile Ser Ala Glu Thr Ala Glu Gln Met Asp Val Asp Gln
785 790 795 800
Gln Glu His Ser Ala Glu Glu Gly Ser Val Cys Asp Pro Pro Pro Ala
805 810 815
Thr Lys Ala Asp Ser Val Asp Val Glu Val Arg Val Pro Glu Asn His
820 825 830
Ala Ser Lys Val Glu Gly Asp Asn Thr Lys Glu Arg Asp Leu Asp Arg
835 840 845
Ala Ser Glu Lys Val Glu Pro Arg Asp Glu Asp Leu Val Val Ala Gln
850 855 860
Gln Ile Asn Ala Gln Arg Pro Glu Pro Gln Ser Asp Asn Asp Ser Ser
865 870 875 880
Ala Thr Cys Ser Ala Asp Glu Asp Val Asp Gly Glu Pro Glu Arg Gln
885 890 895
Arg Met Phe Pro Met Asp Ser Lys Pro Ser Leu Leu Asn Pro Thr Gly
900 905 910
Ser Ile Leu Val Ser Ser Pro Leu Lys Pro Asn Pro Leu Asp Leu Pro
915 920 925
Gln Leu Gln His Arg Ala Ala Val Ile Pro Pro Met Val Ser Cys Thr
930 935 940
Pro Cys Asn Ile Pro Ile Gly Thr Pro Val Ser Gly Tyr Ala Leu Tyr
945 950 955 960
Gln Arg His Ile Lys Ala Met His Glu Ser Ala Leu Leu Glu Glu Gln
965 970 975
Arg Gln Arg Gln Glu Gln Ile Asp Leu Glu Cys Arg Ser Ser Thr Ser
980 985 990
Pro Cys Gly Thr Ser Lys Ser Pro Asn Arg Glu Trp Glu Val Leu Gln
995 1000 1005
Pro Ala Pro His Gln Leu Ile Thr Asn Leu Pro Glu Gly Val Arg Leu
1010 1015 1020
Pro Thr Thr Arg Pro Thr Arg Pro Pro Pro Pro Leu Ile Pro Ser Ser

1025 1030 1035 1040
Lys Thr Thr Val Ala Ser Glu Lys Pro Ser Phe Ile Met Gly Gly Ser
1045 1050 1055
Ile Ser Gln Gly Thr Pro Gly Thr Tyr Leu Thr Ser His Asn Gln Ala
1060 1065 1070
Ser Tyr Thr Gln Glu Thr Pro Lys Pro Ser Val Gly Ser Ile Ser Leu
1075 1080 1085
Gly Leu Pro Arg Gln Gln Glu Ser Ala Lys Ser Ala Thr Leu Pro Tyr
1090 1095 1100
Ile Lys Gln Glu Glu Phe Ser Pro Arg Ser Gln Asn Ser Gln Pro Glu
1105 1110 1115 1120
Gly Leu Leu Val Arg Ala Gln His Glu Gly Val Val Arg Gly Thr Ala
1125 1130 1135
Gly Ala Ile Gln Glu Gly Ser Ile Thr Arg Gly Thr Pro Thr Ser Lys
1140 1145 1150
Ile Ser Val Glu Ser Ile Pro Ser Leu Arg Gly Ser Ile Thr Gln Gly
1155 1160 1165
Thr Pro Ala Leu Pro Gln Thr Gly Ile Pro Thr Glu Ala Leu Val Lys
1170 1175 1180
Gly Ser Ile Ser Arg Met Pro Ile Glu Asp Ser Ser Pro Glu Lys Gly
1185 1190 1195 1200
Arg Glu Glu Ala Ala Ser Lys Gly His Val Ile Tyr Glu Gly Lys Ser
1205 1210 1215
Gly His Ile Leu Ser Tyr Asp Asn Ile Lys Asn Ala Arg Glu Gly Thr
1220 1225 1230
Arg Ser Pro Arg Thr Ala His Glu Ile Ser Leu Lys Arg Ser Tyr Glu
1235 1240 1245
Ser Val Glu Gly Asn Ile Lys Gln Gly Met Ser Met Arg Glu Ser Pro
1250 1255 1260
Val Ser Ala Pro Leu Glu Gly Leu Ile Cys Arg Ala Leu Pro Arg Gly
1265 1270 1275 1280
Ser Pro His Ser Asp Leu Lys Glu Arg Thr Val Leu Ser Gly Ser Ile
1285 1290 1295
Met Gln Gly Thr Pro Arg Ala Thr Thr Glu Ser Phe Glu Asp Gly Leu
1300 1305 1310
Lys Tyr Pro Lys Gln Ile Lys Arg Glu Ser Pro Pro Ile Arg Ala Phe
1315 1320 1325
Glu Gly Ala Ile Thr Lys Gly Lys Pro Tyr Asp Gly Ile Thr Thr Ile
1330 1335 1340
Lys Glu Met Gly Arg Ser Ile His Glu Ile Pro Arg Gln Asp Ile Leu
1345 1350 1355 1360
Thr Gln Glu Ser Arg Lys Thr Pro Glu Val Val Gln Ser Thr Arg Pro
1365 1370 1375
Ile Ile Glu Gly Ser Ile Ser Gln Gly Thr Pro Ile Lys Phe Asp Asn
1380 1385 1390
Asn Ser Gly Gln Ser Ala Ile Lys His Asn Val Lys Ser Leu Ile Thr
1395 1400 1405
Gly Pro Ser Lys Leu Ser Arg Gly Met Pro Pro Leu Glu Ile Val Pro
1410 1415 1420
Glu Asn Ile Lys Val Val Glu Arg Gly Lys Tyr Glu Asp Val Lys Ala
1425 1430 1435 1440
Gly Glu Thr Val Arg Ser Arg His Thr Ser Val Val Ser Ser Gly Pro
1445 1450 1455
Ser Val Leu Arg Ser Thr Leu His Glu Ala Pro Lys Ala Gln Leu Ser
1460 1465 1470
Pro Gly Ile Tyr Asp Asp Thr Ser Ala Arg Arg Thr Pro Val Ser Tyr
1475 1480 1485
Gln Asn Thr Met Ser Arg Gly Ser Pro Met Met Asn Arg Thr Ser Asp
1490 1495 1500
Val Thr Ile Pro Pro Asn Lys Ser Thr Asn His Glu Arg Lys Ser Thr

1505 1510 1515 1520
Leu Thr Pro Thr Gln Arg Glu Ser Ile Pro Ala Lys Ser Pro Val Pro
1525 1530 1535
Gly Val Asp Pro Val Val Ser His Ser Pro Phe Asp Pro His His Arg
1540 1545 1550
Gly Ser Thr Ala Gly Glu Val Tyr Trp Ser His Leu Pro Thr Gln Leu
1555 1560 1565
Asp Pro Ala Met Pro Phe His Arg Ala Leu Asp Pro Ala Ala Ala Ala
1570 1575 1580
Tyr Leu Phe Gln Arg Gln Leu Ser Pro Thr Pro Gly Tyr Pro Ser Gln
1585 1590 1595 1600
Tyr Gln Leu Tyr Ala Met Glu Asn Thr Arg Gln Thr Ile Leu Asn Asp
1605 1610 1615
Tyr Ile Thr Ser Gln Gln Met Gln Val Asn Leu Arg Pro Asp Val Ala
1620 1625 1630
Arg Gly Leu Ser Pro Arg Glu Gln Pro Leu Gly Leu Pro Tyr Pro Ala
1635 1640 1645
Thr Arg Gly Ile Ile Asp Leu Thr Asn Met Pro Pro Thr Ile Leu Val
1650 1655 1660
Pro His Pro Gly Gly Thr Ser Thr Pro Pro Met Asp Arg Ile Thr Tyr
1665 1670 1675 1680
Ile Pro Gly Thr Gln Ile Thr Phe Pro Pro Arg Pro Tyr Asn Ser Ala
1685 1690 1695
Ser Met Ser Pro Gly His Pro Thr His Leu Ala Ala Ala Ala Ser Ala
1700 1705 1710
Glu Arg Glu Arg Glu Arg Glu Arg Glu Lys Glu Arg Glu Arg Glu Arg
1715 1720 1725
Ile Ala Ala Ala Ser Ser Asp Leu Tyr Leu Arg Pro Gly Ser Glu Gln
1730 1735 1740
Pro Gly Arg Pro Gly Ser His Gly Tyr Val Arg Ser Pro Ser Pro Ser
1745 1750 1755 1760
Val Arg Thr Gln Glu Thr Met Leu Gln Gln Arg Pro Ser Val Phe Gln
1765 1770 1775
Gly Thr Asn Gly Thr Ser Val Ile Thr Pro Leu Asp Pro Thr Ala Gln
1780 1785 1790
Leu Arg Ile Met Pro Leu Pro Ala Gly Gly Pro Ser Ile Ser Gln Gly
1795 1800 1805
Leu Pro Ala Ser Arg Tyr Asn Thr Ala Ala Asp Ala Leu Ala Ala Leu
1810 1815 1820
Val Asp Ala Ala Ala Ser Ala Pro Gln Met Asp Val Ser Lys Thr Lys
1825 1830 1835 1840
Glu Ser Lys His Glu Ala Ala Arg Leu Glu Glu Asn Leu Arg Ser Arg
1845 1850 1855
Ser Ala Ala Val Ser Glu Gln Gln Gln Leu Glu Gln Lys Thr Leu Glu
1860 1865 1870
Val Glu Lys Arg Ser Val Gln Cys Leu Tyr Thr Ser Ser Ala Phe Pro
1875 1880 1885
Ser Gly Lys Pro Gln Pro His Ser Ser Val Val Tyr Ser Glu Ala Gly
1890 1895 1900
Lys Asp Lys Gly Pro Pro Pro Lys Ser Arg Tyr Glu Glu Glu Leu Arg
1905 1910 1915 1920
Thr Arg Gly Lys Thr Thr Ile Thr Ala Ala Asn Phe Ile Asp Val Ile
1925 1930 1935
Ile Thr Arg Gln Ile Ala Ser Asp Lys Asp Ala Arg Glu Arg Gly Ser
1940 1945 1950
Gln Ser Ser Asp Ser Ser Ser Ser Leu Ser Ser His Arg Tyr Glu Thr
1955 1960 1965
Pro Ser Asp Ala Ile Glu Val Ile Ser Pro Ala Ser Ser Pro Ala Pro
1970 1975 1980
Pro Gln Glu Lys Leu Gln Thr Tyr Gln Pro Glu Val Val Lys Ala Asn

```
1985                1990                1995                2000
Gln Ala Glu Asn Asp Pro Thr Arg Gln Tyr Glu Gly Pro Leu His His
                2005                2010                2015
Tyr Arg Pro Gln Gln Glu Ser Pro Ser Pro Gln Gln Gln Leu Pro Pro
            2020                2025                2030
Ser Ser Gln Ala Glu Gly Met Gly Gln Val Pro Arg Thr His Arg Leu
        2035                2040                2045
Ile Thr Leu Ala Asp His Ile Cys Gln Ile Ile Thr Gln Asp Phe Ala
    2050                2055                2060
Arg Asn Gln Val Ser Ser Gln Thr Pro Gln Gln Pro Pro Thr Ser Thr
2065                2070                2075                2080
Phe Gln Asn Ser Pro Ser Ala Leu Val Ser Thr Pro Val Arg Thr Lys
                2085                2090                2095
Thr Ser Asn Arg Tyr Ser Pro Glu Ser Gln Ala Gln Ser Val His His
            2100                2105                2110
Gln Arg Pro Gly Ser Arg Val Ser Pro Glu Asn Leu Val Asp Lys Ser
        2115                2120                2125
Arg Gly Ser Arg Pro Gly Lys Ser Pro Glu Arg Ser His Val Ser Ser
    2130                2135                2140
Glu Pro Tyr Glu Pro Ile Ser Pro Pro Gln Val Pro Val Val His Glu
2145                2150                2155                2160
Lys Gln Asp Ser Leu Leu Leu Leu Ser Gln Arg Gly Ala Glu Pro Ala
                2165                2170                2175
Glu Gln Arg Asn Asp Ala Arg Ser Pro Gly Ser Ile Ser Tyr Leu Pro
            2180                2185                2190
Ser Phe Phe Thr Lys Leu Glu Asn Thr Ser Pro Met Val Lys Ser Lys
        2195                2200                2205
Lys Gln Glu Ile Phe Arg Lys Leu Asn Ser Ser Gly Gly Gly Asp Ser
    2210                2215                2220
Asp Met Ala Ala Ala Gln Pro Gly Thr Glu Ile Phe Asn Leu Pro Ala
2225                2230                2235                2240
Val Thr Thr Ser Gly Ser Val Ser Ser Arg Gly His Ser Phe Ala Asp
                2245                2250                2255
Pro Ala Ser Asn Leu Gly Leu Glu Asp Ile Ile Arg Lys Ala Leu Met
            2260                2265                2270
Gly Ser Phe Asp Asp Lys Val Glu Asp His Gly Val Val Met Ser Gln
        2275                2280                2285
Pro Met Gly Val Val Pro Gly Thr Ala Asn Thr Ser Val Val Thr Ser
    2290                2295                2300
Gly Glu Thr Arg Arg Glu Glu Gly Asp Pro Ser Pro His Ser Gly Gly
2305                2310                2315                2320
Val Cys Lys Pro Lys Leu Ile Ser Lys Ser Asn Ser Arg Lys Ser Lys
                2325                2330                2335
Ser Pro Ile Pro Gly Gln Gly Tyr Leu Gly Thr Glu Arg Pro Ser Ser
            2340                2345                2350
Val Ser Ser Val His Ser Glu Gly Asp Tyr His Arg Gln Thr Pro Gly
        2355                2360                2365
Trp Ala Trp Glu Asp Arg Pro Ser Ser Thr Gly Ser Thr Gln Phe Pro
    2370                2375                2380
Tyr Asn Pro Leu Thr Met Arg Met Leu Ser Ser Thr Pro Pro Thr Pro
2385                2390                2395                2400
Ile Ala Cys Ala Pro Ser Ala Val Asn Gln Ala Ala Pro His Gln Gln
                2405                2410                2415
Asn Arg Ile Trp Glu Arg Glu Pro Ala Pro Leu Leu Ser Ala Gln Tyr
            2420                2425                2430
Glu Thr Leu Ser Asp Ser Asp Asp
        2435                2440
```